REPORT RESUMES

ED 003 337 24

TEACHERS' USE OF DICTATING MACHINES TO IMPROVE THE WRITTEN COMPOSITION OF COLLEGE STUDENTS.

BY- KALLSEN, T.J.
STEPHEN F. AUSTIN STATE COLL., NACOGDOCHES, TEX.
REPORT NUMBER CRP-2311 PUB DATE 65
EDRS PRICE MF-$0.18 HC-$4.24 106P.

DESCRIPTORS- SUPERIOR STUDENTS, HIGH SCHOOL STUDENTS, COLLEGE STUDENTS, ENGLISH INSTRUCTION, *COMPOSITION (LITERARY), *ACHIEVEMENT RATING, TEACHER EXPERIENCE, *TEACHER IMPROVEMENT, *TEACHING METHODS, EDUCATIONAL EQUIPMENT, *PHONOTAPE RECORDINGS, STUDENT ATTITUDES, NACOGDOCHES, TEXAS

THE RISING WORKLOAD OF ENGLISH TEACHERS FOSTERED THIS FEASIBILITY STUDY WHICH USED DICTATING MACHINES TO TRANSMIT TEACHER COMMENTS ON COMPOSITIONS TO STUDENTS. APPROXIMATELY 600 STUDENTS IN 30 SECTIONS OF FRESHMAN ENGLISH WROTE PRE- AND POST-EXPERIMENTAL THEMES. IN CONTROL SECTIONS, TEACHERS GRADED ASSIGNED THEMES IN THE TRADITIONAL WAY WITH MARGINAL SYMBOLS AND COMMENTS. IN EXPERIMENTAL SECTIONS, ONLY MINIMUM NOTATIONS WERE ALLOWED ON THE PAPER, AND ANY EXTENDED COMMENTS WERE RECORDED ON PLASTIC DISKS FOR LATER STUDENT USE. IT WAS FOUND THAT 95 PERCENT OF THE TEACHERS AND 65 PERCENT OF THE STUDENTS LIKED THE RECORDS. FOR THE AVERAGE STUDENT, RECORDED COMMENTARY CAN IMPROVE HIS WRITTEN COMPOSITION, BUT THE IMPROVEMENT IS LIKELY TO APPROXIMATE THAT CAUSED BY THE TRADITIONAL METHOD. THE SINGLE, IMPORTANT IMPROVEMENT THAT RECORDED COMMENTARY CAN BRING ABOUT IS IN BETTER CONTENT AND ORGANIZATION, AND THAT IMPROVEMENT SEEMS TO BE RESTRICTED TO SUPERIOR STUDENTS. (JM)

U. S. DEPARTMENT OF HEALTH, EDUCATION AND WELFARE
Office of Education
This document has been reproduced exactly as received from the person or organization originating it. Points of view or opinions stated do not necessarily represent official Office of Education position or policy.

TEACHERS' USE OF DICTATING MACHINES TO IMPROVE THE WRITTEN COMPOSITION OF COLLEGE STUDENTS

Cooperative Research Project No. 2311

by

T. J. Kallsen

Stephen F. Austin State College
Nacogdoches, Texas

1965

The research reported herein was supported
by the Cooperative Research Program
of the Office of Education,
U. S. Department of Health, Education, and Welfare

ED 003337

Preface

This report does not reveal the caliber of the people who shared the hard work of the study. Professor Royal Embree, Jr., of the University of Texas was the consultant. Kinder and more understanding than required, he early recognized my ignorance of statistics and patiently explained necessary procedures and safeguards. Much of the strength of the experiment belongs to his advice; the weaknesses belong to failure in following.

Among my debts is a heavy one I owe to Edwin W. Gaston, Jr., professor of English at Stephen F. Austin State College. As research associate, he supervised the thirty classes participating in the project, set up panels of judges, and helped to develop evaluative standards by conducting reading sessions for teachers--as well as teaching two classes in the experiment. Others on the faculty of the college who also taught two sections and submitted themselves to the discipline of the experiment include Warren B. Austin, professor of English; Gladys Fox, professor of English; Walter E. Kidd, professor of English; Robert L. Weeks, professor of English; Elma Heard, associate professor of English; Myrtle P. Pope, associate professor of English; Sybil Wyatt, associate professor of English; Robert Fusillo, assistant professor of English; Maurice Cecil, assistant professor of English; James R. Parrish, instructor of English and journalism; Joy Rea, instructor of English; Richard Turner, instructor of English; Johnnie Maude Tyler, instructor of English.

Mrs. Pat Fardal suffered the welter of clerical detail with only occasional oral, but pleasant, complaints; and Mrs. Rebecca Camp, who came late into the project, assumed work beyond her normal duties. To Dean J. N. Gerber, who encouraged the experiment; to President Ralph W. Steen, who always supported the project and, noting its probable importance found the money to buy recording machines for future use; to Comptroller Charles Haas, who gave financial guidance through the tangled tape; and to George Hardin, director of the college's Educational Data Processing Center, whose good humor allowed me to face the baleful red lights of monster-IBM with outward equanimity--to all these go my gratitude.

November 1965
Stephen F. Austin State College
Nacogdoches, Texas

T. J. Kallsen
Professor of English
Principal Investigator

CONTENTS

Page

List of Tables v

Chapter I
Introduction 1
The Problem. 1
Objectives of the Study. 2
Related Research 3

Chapter II
The Procedure. 8
The Teachers 8
The Population 8
The Course of Study. 10
Control and Experimental Sections. 11
Uniformity of Sections 11
Administration of the Pre-Experimental Theme 15
Developing an Error-Count Sheet. 16
Developing a Criteria-and-Score Sheet. 18
Determining Panels of Judges 20
Administering the Post-Experimental Tests. 22
Gathering the Data 22

Chapter III
Analysis of Attitudes Toward the Experimental Method 24
Teachers' vs. Students' General Liking 24
Novelty of the Experimental Method 25
Records and Teachers' Professional Attitude. 25
Recorded Comments and Writing or Rewriting Themes. 25
Recorded Comments and Students' Learning 27
Recorded Commentary and Grades 27
Recorded Commentary and the Four Criteria. 28
Recorded Comments versus Written Comments. 30
Developing a Technique for Making Recorded Comments. . . .31
Recorded Comments for Advanced English Classes 34

Chapter IV
Statistical Analysis of Results of Experimental Method . . .36
The Halo of Subjectivity in Theme Evaluation 36
Comparison of Pre-STEP and Post-STEP Results 38
Pre-Theme and Post-Theme Results: COSM. 39
Pre-Theme and Post-Theme Results: Error-Count 41

Chapter V
Conclusion 45
The Basic Question 45
Effect on Content and Organization 45
Effect on STEP Writing Performance 46
Effect on Selected Errors. 46
Basic Recommendations. 46
Secondary Recommendations. 46

Selected References. 48

Appendix Page

A-1: Outline of Study, Regular Sections. 49
A-2: Outline of Study, Superior Sections 53
B-1: Pre-Theme Topics. 59
B-2: Post-Theme Topics 63
B-3: Assignment of Topics to Class Periods 64
C-1: Error-Count Sheet 65
C-2: Criteria-and-Score Sheet. 66
D-1: Detail of Teachers' Scores and Deviations on Eleven Practice Themes. 67
D-2: Summary of Teachers' Deviations from Mean Scores on Eleven Practice Themes. 70
D-3: Fifteen Teachers Classified into Five Panels of Judges by Directional Factor 71
E-1: Students' Reactions to Use of Recorded Comments. 72
E-2: Students' Attitudes Toward Records Categorized 76
F-1: Fifteen Teachers' Reactions to Use of Recorded Comments 79
F-2: Fifteen Teachers' Attitudes Toward Records Categorized 83
G-1: Correlation Coefficients of Judges Paired Within Panels Derived from Pre-Theme Scores and Pre-Error-Counts for Experimental and Control Groups 86
G-2: Mean Scores of STEP 87
H-1: Mean Scores of COSM for All Students. 88
H-2: Mean Scores of COSM for Paired Students 89
H-3: Mean Scores of CO for All Students. 90
H-4: Mean Scores of CO for Paired Students 91
H-5: Mean Scores of SM for All Students. 92
H-6: Mean Scores of SM for Paired Students 93
J-1: Mean Error-Counts (Ten Categories) for All Students. 94
J-2: Mean Error-Counts (Ten Categories) for Paired Students 95
J-3: Mean Error-Counts (Seven Categories) for All Students. 96
J-4: Mean Error-Counts (Seven Categories) for Paired Students 97
J-5: Mean Error-Counts (Three Categories) for All Students. 98
J-6: Mean Error-Counts (Three Categories) for Paired Students 99

TABLES Page

I Teachers' Assignments and Student Loads in Regular Sections. 9

II Teachers' Assignments and Student Loads in Superior Sections 10

III ACT English Scores for Entering Students in Regular Sections. 13

IV ACT English Scores for Entering Students in Superior Sections 14

V Mean Correlations of Judges Paired Within Panels. 37

VI Pre-Theme to Post-Theme Percentage of Improvement, Summary of Appendix H: All Sections. . 39

VII Pre-Theme to Post-Theme Percentage of Improvement, Summary of Appendix H: Regular Sections. 40

VIII Pre-Theme to Post-Theme Percentage of Improvement, Summary of Appendix H: Superior Sections. 41

IX Pre-Theme to Post-Theme Percentage of Improvement on Mean Error-Counts (Ten Categories), Summary of Appendix J 42

X Pre-Theme to Post-Theme Percentage of Improvement on Mean Error-Counts (Seven Categories), Summary of Appendix J 43

XI Pre-Theme to Post-Theme Percentage of Improvement on Mean Error-Counts (Three Categories), Summary of Appendix J 44

Chapter I

INTRODUCTION

The Problem.

During the next decade or more, freshman students will flood college English faculties and strain the traditional practices. The last several years have already given us warning. How to cope with this inundation is a question that colleges must face.

Can we allow the average size of freshman English sections to gradually increase until the theme-grading load of a teacher tests his time, his strength, and his willingness?

Must we more and more rely on inexperienced or unqualified teachers to keep the job of reading themes manageable?

Must we allow each student to write less so that his teacher can be reasonably expected to read and react to the output?

Or, can we assign papers, as usual, but read only part of them?

None of the implied affirmative answers to these questions is as professionally acceptable, I suppose, as saving the good teacher's time in his theme reading and correcting so that he may read more themes than he does now or read the same number more effectively. Although freshman English courses in the United States vary considerably in standards and objectives, one of the universal aims is to get a student to write better as a result of criticism. This criticism the student's teacher usually offers in marginal notations, terminal comments, or personal conference. I do not suggest anything inherently bad in these modes of criticism.

Yet those who use these methods recognize the practical dangers. Under pressure of time, the marginal note may become a cryptic "gr" that refers the student to two hundred pages of a handbook where, presumably, he can locate the information to correct a grammatical error. Or the student reads in a hasty scrawl slanting across several lines of his unhoned prose the single-word criticism "logic!"--complete with towering exclamation point. What the student needs, and what the teacher knows he should provide, is an analysis of how the thinking went awry. The student must get little consolation from the teacher's problem of finding ten or fifteen minutes, or even more, to write in longhand an analysis that is to go into any specifics. Similarly, the terminal comment a teacher knows he should append to a paper degenerates into phrases of meaningless generalizations: "original," "trite," "poor organization," "develop your ideas," "too abstract"; the list is not endless, but it is depressing.

To fight these superficialities, a teacher can, of course, use specific symbols that narrow the student's search in the handbook to a paragraph or two of applicable information; a close acquaintance with the handbook makes this method readily possible and desirable. The teacher can also reconcile himself to the three- or four-hundred word comment the paper needs. I suppose the majority of freshman papers running to six hundred words would be better revised with instructions at least as lengthy as the original paper. If the teacher becomes so reconciled, he must also accept a theme-grading rate of two or three an hour. If he has three classes of twenty-five each with such a paper required only every other week, what he must reconcile himself to is what the janitor in the same building would call two days of work.

The conscientious teacher--and sometimes the Simon Legree co-ordinating the freshman sections insists on the teacher's having a conscience--supplements this individual attention with personal conferences. Somehow, the individualized criticism a teacher puts on a paper hardly ever falls into the category of individualized instruction. My own experience with scheduling personal conferences is that they must be set up for a minimum of fifteen minutes to do much good, and neither I nor my students have been able to adhere to the schedule originally agreed upon.

A combination of the marginal note, the extended commentary, and the personal conference--a combination of these three methods, without adding to the time required of the teacher in grading a paper, ought to be the answer. That, specifically, this study aimed at by trying to find out the effectiveness of cutting a long-play disk (33 1/3 rpm) containing a commentary on the paper. This disk the student would play back at his convenience and after listening to his teacher's criticism would revise his paper.

Objectives of the Study.

The study relied on several assumptions:

First, using arbitrary symbols keyed to a reference or handbook may be efficient and effective. For certain types of errors, that is, the teacher can communicate information to the students reliably and quickly by using the arbitrary symbol if it is keyed to specific errors discussed in a reference book. The reader marks a passage with a symbol in the margin of the student's paper. The student then consults his reference book and revises the passage after digesting the discussion of similar errors. For the simple or mechanical error, in particular, the symbol can be well used.

Second, extended commentaries, sometimes composed of several hundreds of words, are helpful to a student in understanding complex problems. These commentaries may be running commentaries, made as the teacher reads a paper, or they may be terminal, made after the teacher has read a paper and has an over-all view of it. Ideally, the commentary, however desirable, ought to allow the teacher to read a five- or six-hundred word paper by a freshman in twelve-to-fifteen minutes on the average. Therefore, the teacher should be able, roughly, to read the papers of a class of twenty-five students in five or six hours. Or to put it in another way, if a teacher sits down after an early supper--the freshman English teacher rarely has dinner--with the avowed aim of completing an entire set of freshman papers, he will be able to unfold himself from his desk sometime about midnight.

Third, a personal conference complete with chuckles of enjoyment, the overtones of irony, the rasp of sarcasm, and the inflection of commendation often seems desirable at midnight in the teacher's study, at the same moment of the teacher's reading the paper. A plastic disk, inexpensive for the student, approximates many of the conditions: the convenience of talking about the paper while it is still fresh, the teacher's own voice, the freedom from scheduled conferences, the extended commentary without the labor of time-consuming longhand. All this assumes that talking is quicker, more convenient, more intimate, more organic than writing.

These assumptions emphasize the teacher's role in criticizing students' writing. A further assumption is that criticism without the effect of improvement is valueless in education. The basic question the project tried to answer, then, was this:

Will teacher commentary, recorded on a plastic disk at 33 1/3 rpm and made available for replay to students for revision, improve their subsequent, original writing done within three months?

Related, but subsidiary, questions included:

What effect does such commentary have on students' content and organization as opposed to their sentence structure and mechanics?

What effect does such commentary have on selected kinds of errors that students make?

What effect does such commentary have on students' performance on a standardized, objective test of writing ability?

What is the relative effectiveness of the experimental method on above-average students' writing as opposed to average and below-average students' writing?

Related Research.

An unsystematic view of reports on experiments and of opinions about methods is likely to lead the investigator to assume that written composition can be taught successfully in a way limited only by the ingenuity of the teacher.

Kilburn, for instance, implies that too heavy emphasis on teacher criticism is much ado about nothing, that a student who spends fifteen minutes a day of writing on any subject--writing that may be merely checked, glanced at, or read, but not necessarily graded or marked in the traditional formal way--will improve his writing. "I read," says Kilburn, "the serious ones as slowly as their importance warrants, while I simply check the 'doodles' and pass on."[1] The process he suggests is more complex, of course, than the attention-getting quotation I have cited. He outlines four steps in the method he advocates: (1) Students are impressed with the necessity of writing at least fifteen minutes on any subject; quantity itself is important. (2) Next, the student uses his fifteen minutes a day in writing on only one subject, whereas earlier in Step One he could switch from subject to subject within the fifteen-minute period. (3) Next, the student writes on one subject but divides his output into coherent paragraphs. (4) In the last stage, the student is led to practice the more precise details of composition. Finally, of course, teacher criticism must come into this method--either in marginal notes, terminal and running comments, or in personal conference.

Criticism, advice, correction for the student, Laird argues, can be furnished by other students.[2] In an experiment conducted at the University of Oregon, he gathered three or four students together in Monday conferences, where students corrected each other's papers by using handbook symbols and writing a "considered statement" of the good and bad in the paper. At a second conference on Wednesdays, another student gave the paper a second reading and criticism followed by open discussion of papers within the group. With the two written corrections of his paper and the open discussion, the student rewrote the paper for submission to his teacher on the following Monday. The teacher then examined the student criticism as well as the revised paper and adjudicated arguments among the conferees. At the first conference of the week, the teacher put forth a central subject for the week's study: relevance, unity, or whatever.

Laird sees these advantages to the system: (1) his students had more enthusiasm for learning to write; (2) plagiarism ceased; (3) tension between student and teacher lessened; (4) student involvement in learning increased. Laird also listed some dangers of his method: (1) student criticism of papers is likely to end in generalities; (2) student conferences without guidance of a teacher may result in much irresponsible activity, or very little activity; (3) student conferences aggravate the consequences of student irresponsibility, because students are less likely to tolerate the late paper than the teacher, who has built up a resistance to the source of such aggravation; (4) the process requires office and conference space; (5) student correcting may not be well done because of ignorance. Laird's conclusion was that "...the student receives less instruction in writing but more help in learning to write. . ."[3]

---

[1] P. E. Kilburn, "Every Man His Own Pedagogue: A Project in the Teaching of Freshman Composition," The Journal of Higher Education (February 1962), 92.

[2] Charlton Laird, "College English During the Flood," College English (December 1956), 131-138.

Frankly advocating the use of the handbook in correcting student papers is W. A. Clark, who recommends the use of a "diagnostic and evaluation chart" to summarize student errors during a five-minute conference that the teacher holds with the student.[4] Clark's system is especially geared to a laboratory situation of two hours with students conferring at five-minute intervals with the teacher. Clark says that teachers "have no difficulty in reading twelve papers an hour in this way. . . .and that if the teacher cannot discuss the entire paper in the allotted five, a five-minute conference is worth more than fifteen minutes of margin scribbling."[5]

The recommendations of Kilburn, Laird, and Clark are only a few of the multitude that have appeared during the last quarter century, but they indicate a general concern for what the freshman English teacher is to do with all his papers. Perhaps the problem is too overwhelming and complex for solution. Or, perhaps as Lyman said thirty-five years ago, all the trouble comes from the teachers themselves:

> Greatly overburdened as composition teachers have been, largely as a result of their short-sighted methods, they have too frequently conpensated themselves by the easy evaluation of pupils' compositions on the basis of the presence or the absence of mechanical errors. . . .English teachers have been driven more and more into the position of drill masters. . . . As a result, pupils think that subject-matter teachers care nothing about good form and that English teachers care little or nothing about anything else.[6]

The composition teachers' short-sightedness may stem as much from their conscientiousness as from perversity or prejudice. The average composition teacher, I suppose, assumes not only that in the teacher-pupil relation, the experienced or knowledgeable person of the duo should do the reading, the marking, the criticizing, and the grading, but also that students benefit from commentary on their writing--either in marginal notes, terminal and running discussion, or personal interviews.

---

[3] Ibid., 134.

[4] W. A. Clark, Jr., "Neither a Margin Scribbler Nor a Juggler of Numbers," English Journal, College Edition (February 1939), 133-138.

[5] Ibid., 137.

[6] R. L. Lyman, Summary of Investigations Relating to Grammar, Language, and Composition (Chicago: University of Chicago Press, 1929), 188.

Working from these assumptions, recent investigators evidently rely on tradition as much as possible, even while breaking with it. Lay readers sometimes act as the knowledgeable and experienced person to furnish the commentary and marking of a paper. Of course, the lay reader increases the cost of instruction, Ford reporting that the cost annually for high school students is $4.00 a year when the lay reader receives $2.00 an hour.[7] My own guess is that $4.00 will be the cost per student per semester in college freshman English. The great disadvantage is that the teacher, who ought to know his own students' weaknesses, is likely to remain ignorant.

On a cost basis, Lumsden points out that a system of recorded-commentary-with-clerical-transcription is equal to that of the lay reader.[8] But the added cost, Lumsden discovered, reduced the teacher's correcting time by fifty percent when the recorded commentary was transcribed into a typewritten note attached to a student's paper.

Clerical transcription of the teacher's remarks, of course, is more expensive for the school than cutting the commentary on a plastic disk costing about ten cents to the student. Because giving frequent personal interviews with a heavy student-load is almost impossible, Cohen has suggested that individual criticism be given each student by using a voicewriter. In his experiment, the teacher recorded his commentary on a plastic record for use on an Edison Electronic Disc Voicewriter.[9] The student played the record back at a convenient time and revised his paper according to the recorded criticism. This system naturally ties both the student and the teacher to a special machine. If the machine cuts records at variable speeds, the plastic disks cannot be replayed on standard 33 1/3 rpm phonographs--a decided drawback in administering the system.

Most of the studies made with the recording technique testify to the time-saving feature. Those studies I have learned about show that teachers save at least fifty percent of their correcting time. Studies in three high schools--Acalanes High School in Lafayette, California, under Bruce Reeves; Bentley High School in Livonia, Michigan, under William

---

[7] P. M. Ford, "Lay Readers in the High School Composition Program: Some Statistics," English Journal (November 1961), 522-528.

[8] R. Lumsden, "Evanston, Illinois, Township High School Adds to Its Program: The Use of Dictation Machines in Grading English Themes," The Bulletin of the National Association of Secondary School Principals (January 1962), 223-226.

[9] Cohen, "Correcting Compositions Without a Pencil," English Journal (December 1950), 579-580.

Hartman; and Evanston Township High School in Evanston, Illinois, under Robert Lumsden--all report this saving of time.[10] All three schools mentioned used secretaries to transcribe the comments into typewritten notes attached to the paper. Paying the typist $2.00 an hour, at Acalanes High School the transcription cost ran to eight cents a paper. Reeves has experimented with tape recording and bypassing the clerical transcription. He feels that the listening method, as opposed to the typed comments, is inferior. Locating comments directed to a specific student takes too much time and when the tape is erased a permanent record is destroyed. The advantage of the typed comments is that the student has a permanent record on paper that he can study as often as he wishes. The typed comments also provide a record for personal conferences, either with the student himself or with his parents.

At Cubberley High School in Palo Alto, California, Barney Tanner and other teachers cut a plastic disk at 33 1/3 rpm that the student can take home with him to play on his own phonograph.[11] Both students and parents respond favorably to this method and revisions of it have improved both in quantity and in quality. Tanner reports that students "feel cheated if I hand grade a paper when another student gets the disk."[12]

In a brief trial of the recorded-commentary-with-clerical-transcription that I used with a class of twenty-seven college freshmen, every one said that the system was preferable to the method of referring to a handbook with little further commentary by the teacher. All said they learned more. But how reliable is this introspection? How reliable is the subjective evaluation that has been used so far in the studies I have mentioned? If the recorded commentary is looked at favorably by both students and teachers, are these changes of attitude toward learning how to write worth the expense? Should not the subjective reactions be backed with some kind of objective analysis? These are the kinds of questions that initiated this project.

---

[10] Bruce Reeves, "The Dictation Machine as an Aid in Correcting Compositions," Speech made at the annual convention of the National Council of Teachers of English, San Francisco, 23 November 1963.

[11] Ibid.

[12] Ibid.

Chapter II

THE PROCEDURE

The Teachers.

Of the fifteen teachers who participated in the project, nine had the Ph.D. in various areas of language or literature, with one person needing only to complete the dissertation for his Ph.D. and the remaining five holding the M.A. No teacher had less than two years of experience teaching freshman English. Sixty percent of the instructors had at least fifteen years of experience teaching freshman English. Almost three-quarters of the faculty had experience teaching freshman English in at least two different colleges or universities. Over half of the faculty had taught college freshman English in more states than Texas: Massachusetts, Florida, California, Michigan, North Carolina, Georgia, Virginia, West Virginia, Iowa, New York, Wisconsin, Indiana, and Tennessee. In other words, the faculty was experienced, with more than a parochial background.

The Population.

Thirteen of the instructors taught students in Regular Sections, made up of the lower ninety percent of freshman English students as determined by college entrance scores that students submitted to the guidance department of Stephen F. Austin State College. Two other instructors taught Superior Sections, the upper ten percent and declared English majors. The typical enrollment in Regular Sections was about twenty students (Table I); in Superior Sections, about twenty-two (Table II). In other words, at the beginning of the project (30 January 1964) a little more than five hundred students had registered in Regular Sections, with 250 in Control Sections and 257 in Experimental Sections. Eighty-nine had enrolled in the four Superior Sections, with forty-four in Control Sections and forty-five in Experimental Sections. The Regular Sections varied from a low of sixteen students to a high of twenty-two; the Superior Sections, from a low of twenty-one to a high of twenty-three. Persons familiar with freshman English programs will probably agree that these enrollments per section are perhaps below average and that none of the sections was overloaded. I think they will also agree that distribution throughout the sections was rather uniform.

TABLE I

TEACHERS' ASSIGNMENTS AND STUDENT LOADS IN REGULAR SECTIONS

At Beginning of Spring Semester 1964

English 132

	Control Sections		Experimental Sections		Both Sections
Teacher	Section Number	Enrollment	Section Number	Enrollment	Student Load
A	17	19	23	19	38
B	16	19	1	20	39
D	4	20	8	20	40
F	10	21	18	21	42
G	24	18	22	19	37
H	6	20	14	18	38
I	21	18	27	21	39
J	7	20	11	16	36
K	13	19	15	22	41
L	29	18	25	20	38
M	2	20	3	20	40
N	26	20	28	21	41
O	12	18	5	20	38
		Total: 250		Total: 257	Total: 507
		Mean: 19.2		Mean: 19.7	Mean: 39
		Median: 19		Median: 20	Median: 39

TABLE II

TEACHERS' ASSIGNMENTS AND STUDENT LOADS IN SUPERIOR SECTIONS

At Beginning of Spring Semester 1964

English 134

	Control Sections		Experimental Sections		Both Sections
Teacher	Section Number	Enrollment	Section Number	Enrollment	Student Load
C	34	21	31	22	43
E	32	23	33	23	46
	Total:	44	Total:	45	Total: 89
	Mean:	22	Mean:	22.5	Mean: 44.5

The Course of Study.

Students in Superior Sections had passed English 133, a course entitled Composition and Rhetoric: Exposition and Argument, during Fall 1963. Those in Regular Sections during the same time had studied in, and passed, English 131, Composition and Rhetoric: Exposition. Prior to the project (Spring 1964), then, all students had passed one semester of freshman English at Stephen F. Austin State College.

In Regular Sections students studied in a standard program of one year of freshman English devised for students in the lower ninety percent of freshmen at the college. The first semester covers exposition in a course that carries three semester hours of credit and meets either three times a week for fifty minutes each on a Monday-Wednesday-Friday cycle or twice a week for seventy-five minutes each on a Tuesday-Thursday cycle. The second semester of the standard program is a course in argument required of all students who pass the first semester.[1]

In Superior Sections, students studied in an accelerated program of one year that compressed the standard freshman program into the first semester and the standard sophomore literature program into the second semester. The accelerated program is for students in the upper ten percent of entering freshmen and for those who intend to become English majors. The first semester, English 133, carries three semester hours of credit and covers both exposition and argument; classes meet on the same cycles as those in the standard program. The second semester, English 134, covers short stories, novels, drama, and poetry.[2]

---

[1] See Appendix A-1 for an outline of English 131 and 132, the standard, one-year program for Regular Sections.

Control and Experimental Sections.

A total of twenty-six Regular Sections and four Superior Sections participated in the project. Half of each of these groups became Control Sections, with the other half designated as Experimental Sections. Thus, thirteen Control Regular Sections balanced thirteen Experimental Regular Sections (Table I) and two Control Superior Sections balanced two Experimental Superior Sections (Table II).

Two instructors taught the four Superior Sections, each one teaching a Control Superior Section as well as an Experimental Superior Section. Thirteen instructors taught the twenty-six Regular Sections, each instructor being assigned to one Control Regular Section and one Experimental Regular Section. Thus, every instructor assigned to the project taught two sections of freshman English besides his other classes made up principally of literature classes.

The typical student load in the project was just under forty students for a teacher's two Regular Sections, half of whom were registered in Control Regular Sections and half in Experimental Regular Sections (Table I). In contrast, the typical student load of teachers assigned to Superior Sections was about forty-five with about half registered in Control Superior Sections and the other half in Experimental Superior Sections (Table II). The division of students into Regular or Superior Sections occurred before registration for Fall 1963 classes, according to entrance examination scores submitted by incoming freshman students.

Uniformity of Sections.

Although having matched sections as to ability in writing would have been desirable, it was impossible to attain. First, individual schedules required by students for various reasons made shifting them about at will difficult, if not impossible. Second, the rather heavy drop-out rate destroyed any planned matching of sections anyway.

In Fall 1963 students were allowed to register for English 131 and 133 sections as they desired, or were "forced to," until all sections scheduled had about fifteen students enrolled. Then sections were "built up" to twenty students each, next up to twenty-three students, then up to twenty-five, where most English 131 sections levelled off but where English 133 sections continued to build to about thirty students each. Students entered the registration hall at one-hour intervals according to time cards issued by the registrar's office until any one hour would care roughly for five percent of the college's anticipated total enrollment. Presumably, the two methods of distributing students in freshman English sections resulted in satisfactory randomizing. To test this hypothesis, the mean, the standard deviation, and the variance of English scores in the American College Testing Program (ACT) for about sixty percent of the entering students in English 131 and 133 were studied at each hour classes met. The results are presented in Table III and Table IV.

---

[2] See Appendix A-2 for an outline of English 133 and 134, the accelerated, one-year program for Superior Sections.

TABLE III

ACT English Scores for Entering Students in Regular Sections

Fall 1963

Hours Classes Met	Mean Scores	Standard Deviations	Variance
8MWF	14.90	5.54	30.71
9MWF	15.81	4.73	22.44
10MWF	17.40	4.44	19.79
11MWF	15.25	4.37	19.15
12MWF	18.27	4.34	18.88
1MWF	17.50	3.41	11.67
2MWF	17.50	4.14	17.19
7:45TT	16.23	4.32	18.71
9:10TT	16.34	4.47	19.99
10:35TT	16.44	4.73	22.39
12:00TT	17.12	4.18	17.50
1:25TT	16.85	5.10	26.05
2:50TT	17.56	3.73	13.97
5:00TT	17.20	5.45	29.76
All Sections	16.88	4.58	21.02
Median	16.98	4.41	19.47
High	18.27	5.54	30.71
Low	15.25	3.41	11.67

TABLE IV

ACT English Scores for Entering Students in Superior Sections

Fall 1963

Hours Classes Met	Mean Scores	Standard Deviations	Variance
9:00MWF	23.08	3.76	14.16
10:35TT	24.41	2.30	5.31
All Sections	23.68	3.27	10.71

Thus, for example, the first line in Table III reveals that the mean score for English ACT of students enrolled in Fall 1963 at eight o'clock Mondays, Wednesdays, and Fridays was 14.90; the standard deviation, 5.54; and the variance (standard deviation squared), 30.71. Table III also shows that the mean score for all sections of English 131 was 16.88, with the median score almost identical at 16.98. The range between the high mean score (18.27) and the low mean score (15.25) was 3.02. As could be expected because of the preliminary selection, much greater uniformity resulted in sections of English 133, in which English majors and roughly the upper ten percent were placed. Thus, the standard deviations for English 133 sections fluctuated less than those for English 131 sections and were, on the whole, smaller: the standard deviation for English 131 sections was 2.13 but for English 133 sections only 1.46.

Because the sections varied more than had been hoped (from a low variance of 11.67 in English 131 to a high of 30.71), one further standard for assigning students to control or experimental sections was added for Spring 1964 to help match sections as much as possible. No more than two Regular Sections or two Superior Sections were scheduled at one hour and students were distributed through the hours by the same methods used during Fall 1963. Then all students registered at a given hour were arranged in descending order according to scores made on the Writing Test, Form 1A, of the Sequential Tests of Educational Progress (STEP) administered at the end of the Fall 1963 semester; all odd-numbered students were placed in one section meeting at that hour and all even-numbered students in the other section. After that, one of the sections was designated Experimental and the other Control. Although the STEP Writing Test later became one of the measures of growth and therefore ideally should not also have been a part of the procedure of classifying students into sections, this undesirable feature was overlooked to gain the added assurance of better matched sections. Because only sixty percent of the ACT English scores were available and because those scores were nine months old and did not reflect a student's status in January 1964 anyway--the ACT English scores as a principle of classification were rejected. Even so, drop-outs before and during Spring 1964 distorted the matching made before classes met.

Administration of the Pre-Experimental Theme.

After students had been organized into Experimental and Control sections according to procedures just described, all teachers assigned a pre-experimental theme on the second class day in all sections. This "in-class" argumentative paper served the students as preview of the type of writing to be practiced during the semester. For the course itself, the paper served diagnostically to demonstrate the students' current ability in writing and thinking. For the project, all "pre-themes" were stored without marking or comments, later to be mixed with post-experimental themes and submitted to panels of three judges for scoring and detection of errors.

Students wrote the pre-themes under regular classroom conditions that varied, of course, by location within the building and by the hour scheduled. Each student would also write his post-experimental theme under similar circumstances, except for a time-lapse of twelve weeks. In contrast, about two weeks earlier than the beginning of the project, during the last week of Fall 1963, students had taken the pre-STEP Writing Test, Form 1-A, under identical circumstances: at one time in a large auditorium where they had to use lapboards; similarly they took the post-STEP Writing Test, Form 1-B, in the same auditorium, late in the semester of Spring 1964.

Students in both Control and Experimental Sections--whether they met on the fifty-minute, Monday-Wednesday-Friday cycle or on the seventy-five minute, Tuesday-Thursday cycle--had fifty minutes to write an argumentative paper of at least four hundred words, without consulting notes or references like handbooks or dictionaries.[3] Although the use of references can be advocated because it makes the writing situation more realistic, for the project students were denied this privilege to assure uniformity of conditions: obviously, some students were going to appear for the pre-theme without reference books because of carelessness or neglect or poor memories. The purpose of the project was not to study the effect that a student's acceptance of responsibility had on his writing.

Each student had a set of four questions,[4] one of which he chose to argue. Each set of four questions contained two of national significance and two of local collegiate significance to give each student "something to write about." Teachers assigned the same set during a pair of consecutive periods on the theory that students meeting during the first class period could not easily pass on their list of questions to students who would write on them in the following period (only ten minutes intervened between classes).[5]

---

[3] See Appendix B.

[4] See Appendix B-1.

Developing an Error-Count Sheet.

After the pre-themes had been written during the second class period, teachers read them without writing comments or grades on the papers. Although this method of reading was not normal for the course, I felt that each paper had to be clean of any teacher's reactions when the panels chosen for the project came to judge the papers as examples of end-of-semester arguments. Teachers, then, did not return pre-themes for student revision as they ordinarily would have if the pre-themes had served their usual purpose as diagnostic papers. Instead of being revised, the pre-themes went into a central file where they were stored until three months later when the post-experimental themes were written and both the pre- and post-themes were to be read by a panel of three judges.

Teachers went ahead in their classrooms teaching their Control and Experimental Sections similarly according to the outline of study regarded as standard for the course (see Appendix A). Teachers, of course, would require student-revision of all so-called teaching themes: in the Control Sections all suggestions for revision came in the traditional manner through marginal notation and written terminal comments; in the Experimental Sections, however, only minimum written notations were used and teachers substituted recorded comments for the traditional truncated written comments.

During the first half of the project, outside the classroom, teachers developed criteria and definitions of errors to guide them when they became judges for evaluating pre- and post-themes. For the faculty itself this period of study, discussion, and even wrangling may have resulted in a by-product more valuable than the answers to the central questions of the project.

A first decision in developing a check list of errors was never changed; namely, that the list of errors had to be selective, not exhaustive. A second assumption was never changed either: that the errors finally selected should be those that could be rather easily defined and objectively detected.

The first trial list of errors used in practice was too specific, too narrow, too selective. Items that directed the teacher's attention, say, to the student's failure to place a comma between the city and the state in an address were unlikely to appear frequently enough in the free writing allowed by the project. Errors, in other words, had to be defined so that likelihood of occurrence would lend some statistical validity to the study. Simultaneously, of course, loosening the definition made the task of detection more imprecise. I suppose none of the fifteen teachers was finally satisfied with the list of errors decided upon. Even though the teachers might agree upon a definition, writing is so personal that time after time students wrote sentences with constructions on the borderline of an already accepted definition.

---

[5] See Appendix B-3.

Thus, for example, the final error-count sheet developed for use in the project had one category of the ten chosen called "sentence sense," defined as covering "fragments, comma splices, fused sentences."[6] These last three terms are common enough with professional teachers of writing, but they are, nevertheless, incapable of complete objective application. The term "fragment" implies a definition of the term "sentence" and the term "sentence" has had literally dozens of definitions. We decided arbitrarily that requiring a sentence to have a subject and predicate was too confining because, for instance, it would label the "fragment" one often uses to answer a question as an error. In the same way, we accepted the comma as a proper separator between independent clauses without coordinate conjunctions when the sense indicated close contact of ideas, as in the now-famous example "I came, I saw, I conquered." Even so, borderline cases remained. Whether the following sentence illustrates the error called a comma splice will start an argument: "Women should have the same rights as men at Stephen F. Austin State College, they shouldn't be required to eat at the cafeteria." Simple examples like this quickly exposed the moral, philosophical, linguistic, and prejudicial nerve-ends of individual teachers. And I do not imply that the fifteen teachers trying to arrive at definitions were any more sensitive or irrational than the average teacher in the profession. I think, instead, our discussions were representative in uncovering the problems that challenge, if not plague, the teachers of written composition.

In any case, the faculty of fifteen had stepped into the familiar reiteration of the basic, unsolved problems--and we did not have the time for such an analysis, even assuming immodestly that we could reach absolute agreement. Under the handicaps of time and objectives of the project, we ultimately had to assume understanding of definitions. The inadequacy of definition, coupled with the human frailty of overlooking an error, resulted in diversity of reporting. Undoubtedly an English teacher should detect the misspelling in "the principle objective"; but if he overlooks it for whatever reason, he can only apologize with a red face at being caught. Under the usual circumstance of reading themes, then, even the reporting of misspellings--perhaps the most obvious and most easily defined error--is not accurate.

After the discussions and trials with different kinds of check lists, a final guide to error-count with ten categories emerged: (1) sentence sense: fragments, comma splices, fused sentences; (2) agreement: subject-verb and pronoun-antecedent concord; (3) reference: vagueness, ambiguity, indefiniteness; (4) verbs: tense and principal parts; (5) case: nominative, objective, possessive; (6) modifiers: comparative and superlative forms, danglers; (7) parallelism: illogical construction in lists and comparisons; (8) spelling: unconventional variants; (9) capitalization: unconventional particular cases; (10) punctuation: unconventional formal punctuation like omitting periods at ends of sentences or with abbreviations; omitting commas in dates, addresses, series, non-restrictives, mild exclamations; omitting hyphens in compounds; omitting quotation marks in direct quotations and titles. An actual error-count sheet itself is reproduced in Appendix C-1, with directions to the teacher for using it.

---

[6] See Appendix C-1.

Developing a Criteria-and-Score Sheet.

If defining errors and then detecting them under the definitions were difficult, even more so were defining criteria and applying them to derive a score representing the quality of a paper.

One criterion we never changed from beginning to end of our considerations: that on a percentage basis the evaluation of a paper should reflect roughly sixty percent emphasis on content and forty percent on presentation. One other standard, a quantitative one, referred to the minimum length of a paper written in fifty minutes; past experience with our students led us to think that a freshman could be expected to write at least four hundred words in a fifty-minute period.

With these assumptions, then, we tried first of all a simple five-point scale reflecting the traditional A-B-C-D-F grades so familiar to collegiate grading systems. But one trial with a practice theme assured us that such a scale was too gross to reflect what we liked to think of as our ability to make more precise judgments. Although the gross scale would undoubtedly enable us to report much more uniformity--at least seeming uniformity--we quickly decided that the scale we adopted would need more divisions to show measurable changes in a student's performance after a lapse of twelve weeks.

In accordance with this conclusion, we next tried a fifteen-point scale with twelve points allotted to gradations in passing work and three points allotted to failing work. The fifteen point scale obviously was a mere refinement of the five-letter scale already referred to. Thus, A+ was worth fifteen points; A, fourteen; A-, thirteen; and so on down to the F category with F+ earning three points; F, two points; F-, one point. Applying this scale we found it to be more satisfactory because it allowed us figures that expressed more precision of judgment. We also were growing more dissatisfied with the sixty percent to forty percent emphasis between content and presentation. We therefore hoped to get greater precision on the fifteen point scale by breaking content into two categories of equal importance (content with thirty percent emphasis and organization with thirty percent); the other forty percent emphasis we also broke into categories of equal importance (sentences with twenty percent emphasis and mechanics with twenty percent). This categorization resulted in some lapping over between categories, as, for example, between content and sentences. Thus a student's generalizations and abstractions were likely to be expressed in clichés or in general and abstract terminology (located in our categorization under the heading of "sentences"). In spite of this handicap, we decided that the four main categories with their relative emphases could be applied with a little care and alertness.

But the three points on the failing scale when distributed across the four main categories of content, orgainzation, sentences, and mechanics did not carry enough weight. Too many teachers complained that the scoring system resulted in passing grades for papers that the teacher's overall judgment of the theme could not sustain. In other words, we thought we wanted a scale weighted more heavily on the failing side in each of the four main categories. We next tried a variation of the five-point scale by expanding it to fifty points with ten points representing the gradations of judgment within each letter of the old five-letter scale. Applying this scale to two more practice themes revealed quickly it was no more manageable than the fifteen-point scale. Ten points on the failing scale could no more balance forty points on the passing scale than the old three points on the failing scale could balance twelve on the passing.

At this juncture, we reasoned that a scale should logically have as many points on the failing scale as it had on the passing scale--so we tried eighty points: forty points to passing and forty points to failing. This, of course, was only another variation of the five-letter scale, with the top forty points representing the *A-B-C-D* of passing work and the bottom forty representing the *F* of failing work. So far, this was the best scale we had developed, but we discovered that we had not accustomed ourselves to think of gradations in failing although we were accustomed to rather precise gradations of passing work. We could judge whether a particular paper failed or passed with considerable harmony, but we argued about how badly a paper might fail more than we argued about how well a paper might pass. And in our training sessions we never did overcome this weakness. Our greatest discrepancies of judgment throughout the training sessions came when we were confronted with the failing paper. Even after we had completed our formal training sessions I suppose that we had not erased this relative discrepancy.

Obviously, as judges we were victims of years of assumptions about grading that could not be eliminated in a few weeks of even well-intentioned work. After experience with the eighty-point grading scale, someone suggested that since the eighty-point scale was so close to the familiar one-hundred-point scale , why should we not go back to it? With some relief, we did. So the top forty points were allotted equally to *A-B-C-D* and the bottom sixty points to *F*.

The final "Criteria-and-Score Sheet for Rating Themes," reproduced in Appendix C-2, had four main categories: (1) content, worth a total of thirty points for a "perfect" score; (2) organization, also thirty points; (3) sentences, twenty points; and (4) mechanics, twenty points. The divisions, of course, were arbitrary and based on relative emphases developed during the teaching of the courses in the preceding five or six years.

The category of content referred to relevance, proportion, specificity, and concreteness of material. Organization referred to effectiveness of introduction, logical ordering of main divisions, coherence through transitions, logical development of subdivisions, and effectiveness of the conclusion. The category of sentences was a portmanteau that contained ineffective fragments, comma splices and fused sentences; lack of clarity from misplaced or dangling modifiers, split or nonparallel constructions, ambiguous reference, shifts in viewpoint or tense, verbiage, over-elegance, mixed figures of speech; ineffective passive voice, triteness; inappropriate tenses, principal parts of verb, inflections; lack of agreement between subject-verb or pronoun-antecedent. Mechanics--as a category less comprehensive and more clearly cut--referred to punctuation that muddled communication, broke the conventions of the style book, drew the reader's attention to its vagaries; the category further included misspelling and non-standard capitalization.

Determining Panels of Judges.

After the faculty had developed the final Criteria-and-Score Sheet for Rating Themes by applying proposed standards to nine practice themes, an additional eleven themes graded by the Criteria-and-Score Sheet prepared the faculty for judging pre- and post-experimental themes.

These eleven practice themes were assigned for rating, two or three at a time, over a period of about three weeks. At a discussion period, after a set of two or three themes had been rated, each teacher learned how far he deviated from the mean rating for a theme (Appendix D-1 lists the results). In addition, teachers discussed scores and standards after each theme-reading. If a teacher differed radically from the mean score, he tried to make a conscious adjustment when the next set of practice themes was assigned. That this system had some degree of success can be inferred by comparing the average deviation of ratings for the first seven themes with the average deviation for the last four.

Apparently, on the first seven practice themes teachers were busy making adjustments to the Criteria-and-Score Sheet, because the average deviation from the mean rating ranged from a low of 4.9 to a high of 9.0 on the first seven themes, with a mean average deviation of 7.7. In contrast, the average deviation on the last four practice themes ranged from a low of 5.5 to a high of 7.3--with a mean average deviation of 6.0.

Considering only those themes on which the mean score indicated a "passing" grade shows that teachers were obviously in better agreement on the last four themes than on the first seven:

	Theme 2	Theme 3	Theme 6	Theme 8	Theme 9	Theme 10
Mean Score	67.8	78.6	76.3	82.7	60.7	78.1
Average Deviation	4.9	7.5	7.7	5.7	5.5	5.6

Thus, the mean average deviation for Themes 2, 3, and 6 was 6.7; but for Themes 8, 9, and 10 (from the last four) it was 5.6.

But if one considers only those themes on which the mean score indicated a "failing" grade (below sixty) teachers showed less harmony in their judgments:

	Theme 1	Theme 4	Theme 5	Theme 7	Theme 11
Mean Score	39.3	51.3	36.6	55.6	55.1
Average Deviation	8.3	9.4	6.9	9.0	7.3

Thus, the mean average deviation for these five failing themes was 8.2; the mean average deviation for failing themes among the first seven practice themes was 8.4 and for the only failing theme in the last four practice themes, 7.3. Presumably, more practice on sample themes would have lowered the average deviation still more, both on passing and on failing themes. In any case, the growth toward harmonious judgment shown in the last four themes appeared to be satisfactory training for rating the experimental themes.

After the practice in using the Error-Count Sheet and the Criteria-and-Score Sheet on the eleven sample themes, the deviation of each teacher from the mean score of all teachers on given themes was rounded off (as listed in Appendix D-2). Then the algebraic sum of each teacher's deviation from the mean score on all eleven themes was computed to derive a "directional factor," which became the basis for assigning teachers to panels of three judges each. Thus, for example, the directional factor for Teacher A was -55, which shows that his scores were on the average five points below the mean on each of the eleven practice themes. Teacher F, on the other hand, had a directional factor of +84, indicating that his scores were on the average almost eight points higher than the mean score. In other words, the directional factor for Teacher A would probably predict that his grading of a particular theme would be lower than the "average" teacher's grading and that Teacher F's grading would probably be higher than the average teacher's. At least, this assumption became the basis for setting up five panels of judges.

First, we divided the fifteen teachers into three groups of five each (as detailed in Appendix D-3). One group of five, the "lowest one-third" in the minus direction, had the five highest minus directional factors of the fifteen--ranging from -23 to -96. Another group of five, the "highest one-third" in the plus direction, had the five highest plus directional factors of the fifteen ranging from +20 to +84. The third group of five, the "middle one-third" in both plus and minus directions, were the "average" graders on the faculty with directional factors ranging from -10 to +19.

After this preliminary classification into "high," "middle," and "low" graders, a final classification into five panels resulted by juggling the three teachers in each group until the cumulative directional factor for each panel was as similar as possible. As Appendix D-3 shows, the five panels had a mean cumulative direction of +10.2, with the cumulative direction varying from a low of +7 to a high of +14. The average deviation of the five panels from the mean cumulative direction was 2.64. Finally, then, the mean cumulative direction for each panel of three judges was only +.92 points. This system gave us some assurance that the panels were reasonably balanced by distributing the five so-called "hard" graders, the five "easy" graders, and the five "average" graders one to a panel. The five panels, consequently, were made up as follows: Panel 1, with Teachers D, F, and C; Panel 2, with Teachers A, E, and K; Panel 3, with Teachers I, G, and N; Panel 4, with Teachers L, J, and H; and Panel 5, with Teachers O, B, and M.

Administering the Post-Experimental Tests.

After twelve weeks, during the week of 30 April-6 May 1964, all students remaining in the classes not only wrote their post-theme based on the same sets of topics rearranged as listed in Appendixes B-2 and B-3 but also took the Writing Test Form 1B of the Sequential Tests of Educational Progress. In both tests the conditions were similar: (1) students wrote their themes in fifty-minute periods as impromptu exercises in typical classroom situations without reference materials like dictionaries or handbooks, and (2) they took the STEP Writing Test in a large group gathered in a single auditorium with lapboards to write on. Students who missed the post-theme or Form 1B of the Writing Test because of sickness or other excusable reasons were given the opportunity to "make up" their absences individually.

Gathering the Data.

After a teacher (1) had graded the STEP Writing Test Form 1B with a standard scoring key and (2) had read the post-theme without written comments on the paper itself (both to derive a final examination grade for the course), he returned both the Writing Test answer sheets and the post-theme to the general file.

First, students' names and other identifying marks on both the pre- and post-themes were obliterated and replaced with a code number representing the name of the student who wrote the theme and whether or not the paper was a pre- or post-theme.

Next, both the pre- and post-themes--still kept in class groups--were classified into five batches corresponding to the five panels of judges. The themes were so classified that no member of a panel would read any papers written by a student in one of his classes. Thus each panel would read the pre- and post-themes of six classes, three experimental and three control sections, none of which a panel member had taught himself.

Then, all pre-themes and all post-themes assigned to a panel were mixed so that no judge would find a clue as to whether any particular paper was either a pre- or a post-theme. After this mixing, each batch of papers for a panel was redivided into six approximately equal sub-batches, three of which were immediately issued to the judges on that panel with the other three sub-batches held in reserve to even the "flow" of papers from one judge in a panel to another. Each judge, finally, read all the papers in the six sub-batches and filled out for each theme an Error-Count Sheet (see Appendix C-1) as well as a Criteria-and-Score Sheet (see Appendix C-2).

Each student, therefore, had three sets of scores equivalent to grades determined by three different judges, none of whom had been his teacher during the semester, on both the pre- and post-theme: these scores were broken into four parts: (1) content, (2) organization, (3) sentences, and (4) mechanics. Each student also had three sets of error-counts made by the same three judges under ten categories: (1) sentence sense, (2) agreement, (3) reference, (4) verbs, (5) case, (6) modifiers, (7) parallelism, (8) spelling, (9) capitalization, and (10) punctuation. Each student also had a score for a pre-experimental writing test, the STEP Writing Test Form 1A, and for a post-experimental writing test, the STEP Writing Test Form 1B. These scores became the basis for our primary statistical study.

In addition to these procedures to yield objective measures for study, a questionnaire distributed to both faculty and students in the eighth week of the experiment yielded responses by which attitudes toward the experiment might be determined. These results appear in the next chapter.

Chapter III

ANALYSIS OF ATTITUDES TOWARD THE EXPERIMENTAL METHOD

Although trying something new is likely to arouse feeling pro or con and although liking or disliking a procedure says nothing about its effectiveness, good or bad--nevertheless, a questionnaire of twenty-four items submitted to students in the experimental sections[1] and one of twenty-eight items submitted to fifteen teachers who participated in the experiment[2] yielded information about the subjective success or failure of the experiment.

Teachers' vs. Students' General Liking.

Teachers liked using the records more unreservedly than the students. Almost two-thirds of the students reported favorably that they liked the records from the beginning to the end of the experiment, but one-fifth never did like the records.[3] In contrast, only one of the fifteen instructors said that he never liked the records.[4] I can guess at the reason: some students, who did not own their own record players, had to use those installed in the library or in the English Department; they had to go to extra trouble to listen to the comments on their themes. Further, some students with their own record players ran into technical trouble: some were bothered by too little volume, some reported that their record players rejected the tone arm too early so that part of the commentary was omitted toward the end of the recording, still others reported that a too-light tone-arm skipped some of the grooves.

Teachers, on the other hand, had exclusive use of a portable Soundscriber with its own power-pack. A higher quality recording machine--one not so easily portable and without a self-contained power supply, but one that cuts a bigger record to eliminate too-early rejection of the automatic tone arm and that cuts slightly deeper grooves with more volume--is available from the same manufacturer for a cost of about sixty percent more.

---

[1] Appendix E-1.

[2] Appendix F-2.

[3] Appendix E-2b.

[4] Appendix F-2b.

Novelty of the Experimental Method.

Trying something new is also likely to present technical problems beyond those I have mentioned. For example, would the novelty of the records increase or decrease their effectiveness? Students divided themselves rather evenly when they considered this question: about one-third of the students felt that the records were more effective at the beginning of the semester and an almost equal number thought they were more effective toward the end of the experiment.[5] Over half the teachers thought their commentaries improved during the progress of the experiment if one takes the sum of "yes" and "perhaps so" answers,[6] but on a cross-check item[7] almost half of them thought their comments were as good at the beginning as later. The conclusion seems to be that the newness of the experiment had little to do with how effective either students or teachers thought the records were.

Records and Teachers' Professional Attitude.

Whether the novelty of the records dispelled any prejudice teachers might have harbored toward the machines, I have no evidence for; but any fear of the machines as a dehumanizing influence on education was wiped out. Thirteen out of the fifteen teachers said that they did not regard the experimental use of dictation machines as dehumanizing.[8] Only one admitted that cutting a record struck him as a bit unprofessional and gimmicky.[9] Two reported some compulsion to talk when they faced the blank record, some obligation, as it were, to give the student his money's worth--even though the records did not cost him a penny; two others also admitted that they occasionally felt a similar compulsion.[10]

Recorded Comments and Writing or Rewriting Themes.

That recording comments is not wholly unprofessional is further borne out by both teachers' and students' replies to items asking about the effectiveness of the disks for guiding revision. The students thought more of the disks than their teachers did. A full two-thirds of the students said that recorded comments eased their task of revision and only one-sixth that the records confused them.[11] Teachers, however, were less complimentary, only two-fifths believing that recorded comments resulted in superior revisions when their experimental sections were compared with their control classes.[12] Further, from the teachers' viewpoint, recorded comments evidently made the job of checking revisions harder; at least half the teachers thought so.[13]

---

[5] Appendix E-2b.

[6] Appendix F-1, Item 28.

[7] *Ibid.*, Item 22.

[8] *Ibid.*, Item 1.

[9] *Ibid.*, Item 15.

[10] *Ibid.*, Item 24.

In any case, however, three-quarters of the teachers felt they were aiding their students more with the recorded commentary than with the traditional way of marking papers,[14] but about one-half of them also felt that cutting a record takes too much of the teacher's time as weighed against the value students receive from such help.[15] By rough estimation, a teacher took about twice as much time cutting a record while he was grading a paper, so that the time required for reading papers approximately doubled in the experimental sections. This extra time undoubtedly accounts for the teachers' contradictory feelings that the recorded commentaries were superior but not worth the effort. This same extra effort, however, probably led to the students' feelings that the records were worthwhile. How the relation between these contradictions can be objectively assessed I do not know. After all, how much of a teacher's effort should be required for how much of the student's benefit?

The students obviously felt their disks had value because two-thirds of them said the recorded comments helped when they were assigned a new paper to write.[16] Perhaps as many would say that the traditional method was equally helpful; unfortunately I have no cross-check. Almost four-fifths of the students said they replayed records to double-check their mistakes, and over one-third said they played records more than once during the twelve weeks to see whether they continued making the same mistakes and receiving the same comments on successive papers.[17] Only one-fifth of the students thought teachers took too much time in their commentaries arguing with their ideas and slighting sentence structure, usage, and the like.[18]

---

[11] See Appendix E-2c.

[12] See Appendix F-2c.

[13] Ibid.

[14] Appendix F-2d.

[15] Ibid.

[16] Appendix E-2d.

[17] Appendix E-2c.

All in all, then, students probably felt that recorded comments were more valuable as guidance in original writing and in revision than their teachers did. Perhaps the teachers' attitudes were too much affected by the amount of extra time they knew they had to put into recorded commentary. Of course, I have pointed out that in the traditional method the student did not get the volume of comments that he received in the experimental method. Students were undoubtedly sensitive to this difference in quantity. I suggest that they probably never before had had the feeling of so much personal attention directed to their written composition.

Recorded Comments and Students' Learning.

The generally favorable attitude of students toward recorded comments is borne out by other preferences they expressed. Almost two-thirds of the students in the experimental sections, for example, thought they "got more out of" freshman English because records were used to comment on their themes, and they wished that they could have had the method used during their first semester of freshman English; not only that, they wanted the records to be used in future English classes they had to take.[19] Slightly more than one-fourth of the students disagreed. This one-fourth may include many of the same students who said, in comparing the experimental method with the traditional method of the preceding semester, that they did not learn more with recorded commentary.

Two-fifths of the students, however, were equally sure that they had learned more about how to write a paper with the recorded commentaries, and an additional one-sixth thought perhaps they might have learned more.[20] Over one-half the students, therefore, thought that, in some degree, recorded commentary taught them more about how to write; in other words, in a class of twenty students, eleven thought recorded commentary was superior to the traditional method, five disagreed, and four were indifferent.

Recorded Commentary and Grades.

Although over half the students said they thought they got more out of freshman English with the recorded commentaries, only about a third thought they may have received higher grades because of the records.[21] To a certain degree, the instructors agreed: three of the fifteen teachers (twenty percent) thought their grades might have been higher than otherwise, but twelve of them thought they had not raised grades; in fact, eleven of those twelve said they thought their grades were unaffected.[22]

---

[18] Appendix E-2d.

[19] Appendix E-2e.

[20] Appendix E-1, Item 22.

[21] Appendix E-2e.

The respondents, however, may have found the questions used for eliciting this information to be ambiguous. Thus, it is not clear now whether the students' answers meant that they thought the records were a cause leading them to better performance or whether they thought their teachers were more lenient in their grading as a concomitant of the method. The response of the instructors more likely meant leniency on their part--a leniency induced by the method.

Recorded Commentary and the Four Criteria for Grading.

The grades awarded to themes written during the experiment were determined by applying standards classified under four categories: (1) content, thirty percent; (2) organization, thirty percent; (3) sentences, twenty percent; (4) mechanics, twenty percent.

In which of these four categories were the recorded comments most helpful to students? Teachers and students did not agree when they responded with a "yes," "no," "I don't know," "perhaps so," or "probably not" to statements like

> Of the four categories--content, organization, sentences, mechanics--I think the recorded comments helped me/students most to improve my/their content.

For example, no teacher answered "yes" to such a statement referring to the improvement of mechanics,[23] but one-fourth of the students said, "Yes, recorded comments helped me to improve my mechanics most."[24] For teachers, the "yeses" were divided evenly among the other three categories, twenty-seven percent for each.[25] But thirty-one percent of the students said "yes" for content,[26] twenty-seven percent for organization,[27] twenty percent for sentences.[28] The greatest margin of disagreement, of course, lay in the category of mechanics, twenty-five percent "yeses" for students as contrasted with none for the teachers.

---

[22] Appendix F-2e.

[23] Appendix F-1, Item 13.

[24] Appendix E-1, Item 15.

[25] Appendix F-1, Items 3, 6, 9.

[26] Appendix E-1, Item 3.

[27] *Ibid.*, Item 7.

[28] *Ibid.*, Item 11.

Because teachers graded papers for mechanical errors by marking symbols in the margins of a student's paper to refer him to a handbook explanation, students may have confused this help with the entire experimental method. Of course, a teacher also recorded advice about mechanical errors when he found it desirable; perhaps students appreciated these rather infrequent comments as more valuable than the more inconvenient reference to the handbook. In any case, it is not clear why one-fourth of the students thought recorded commentary helped them most to improve their mechanics, and none of their teachers did.

Combining the "yes" responses and the "perhaps" responses as "favorable" replies to the information sought reveals that sixty percent of the teachers thought their recorded comments helped students most in improving their organization; forty-seven percent were "favorable" instead to content; forty percent were favorable to sentences; and only seven percent were favorable to mechanics.[29] Students' favorable answers in the same categories were forty-nine percent for organization, fifty-four percent for content, forty-one percent for sentences, forty-seven percent for mechanics.[30] Students were evidently quite unclear about how the recordings helped them most.

I suspect that the reason for this confusion lies with students' inability to distinguish sentences, for example, as a separate category from mechanics. Similarly, they probably confused content and organization as categories not meant to overlap. Unlike the students, the teachers spent about eight weeks, in one or two discussion periods a week for as much as two hours at a time, defining the four categories and trying to distinguish them one from another. Students' responses to these categories, therefore, are probably less valid than those of the teachers'.

Because of my early suspicion that students might not be able to clearly distinguish content from organization or sentences from mechanics, I added two other categories to the questionnaire: (1) content and organization and (2) sentences and mechanics--two categories that reclassify the four into two more easily differentiated groups. The responses to these two "catch-all" categories probably make more sense. Thus, forty-seven percent of the teachers said "yes" to the category of content and organization,[31] as did fifty percent of the students.[32]

---

[29] Appendix F-2a.

[30] Appendix E-2a.

[31] Appendix F-1, Item 18.

[32] Appendix E-1, Item 19.

Adding the "perhaps so's" to the "yeses" to form a "favorable" response reveals similar agreement: sixty percent of the teachers favorably responded to the question of whether recorded comments helped their students most to improve their content and organization;[33] the same item raised the percentage of students to fifty-six percent favorable replies.[34]

In the second category of sentences and mechanics, thirteen percent of the teachers[35] and twenty-nine percent of the students[36] said "yes." Again, adding the "perhapses" to the "yeses" showed that twenty percent of the teachers gave "favorable" responses to this item[37] as compared to forty percent of the students.[38]

Considering both the students' and the teachers' replies, we can only conclude that both thought recorded commentary improved students' content and organization more than their sentences and mechanics. Teachers were surer of this than their students--probably because teachers may have been guilty of unconscious self-fulfillment of a prophecy they had made before the experiment started. I think the teachers were less confused than the students; but perhaps the teachers, more fully aware of the implications and structure of the experiment, were more prejudiced.

Recorded Comments versus Written Comments.

Surprisingly, I think, over one-half the students (fifty-three percent) would choose the recorded comment over the same comments written out and attached to the paper.[39] Evidently the spoken word carries value that an English professor may lose track of in his devotion to the written word. Less surprising was the reaction of nearly two-thirds of the students (sixty-two percent) who would choose a personal conference with the teacher to a recorded commentary. Throughout the experiment a common remark I heard among the teachers was that the recorded commentary could never replace the personal conference. The students agreed. One instructor in particular voiced the fear that the plastic disk would become a substitute for the personal conference:

---

[33] Appendix F-2a.

[34] Appendix E-2a.

[35] Appendix F-1, Item 25.

[36] Appendix E-1, Item 24.

[37] Appendix F-2a.

[38] Appendix E-2a.

[39] Appendix E-2e.

> I find myself (she reported after a trial of a month with recorded comments) conversing with the student, via the record, rather than making corrections. Ordinarily, I reserve the conversations for personal conferences. Admittedly, the recorded comment takes the place to a certain extent of the longer conference, but I have the fear that I have possibly offended or confused the student in a way that could have been promptly corrected in the give-and-take of a face-to-face conference. One of my criticisms is that the record is a one-sided discussion. The teacher cannot tell whether he is getting through to the student without making another comment, written or oral. I admit that his uneasiness is probably a temporary adjustment to a new device.

The uneasiness is probably also a genuine teacher's concern for really communicating, not just spooning medicine out regardless of the taker's attitude.

Developing a Technique for Making Recorded Comments.

That some adjustment to cutting the records is necessary several of the teachers admitted. Anyone first trying the method, then, must probably develop some procedure for efficiency--even though it cannot be a common one for all teachers. After six weeks of using the recorded commentary, one instructor reported,

> I discovered that the attempt to use the disk for a running commentary was futile. The mere physical problem of handling pen, paper, handbook, microphone, record book, and smoking materials was insurmountable. I might have become more dexterous with practice; however, I was averaging over an hour per paper and felt that the comments recorded in this way were very poor. They consisted largely of spontaneous reactions to sentence errors that could be adequately treated with handbook symbols to which I had systematically addicted myself last semester. I abandoned the running commentary six or seven papers after I began it. I did not, perhaps, give the method a fair trial; at the rate I was going I didn't think I could afford to.

In contrast, another teacher adopted the running commentary at the beginning of the experiment and continued to use it throughout the semester.

> I liked to interrupt my reading of a student's paper from time to time to let him know how I was reacting, perhaps by telling him what was going through my head and even guessing in what direction

> his logic ought to take him. This method of immediate reaction, of course, led to occasional improper inferences on my part. When that became apparent to me, I again told the student so--probably complimenting him at that time for not having fallen into a trap that he seemed to be setting for himself, or probably scolding him for illogicality, and then informing him what he needed in the way of revision. Outside of the personal conference, I know of no other way of letting the student in on a perceptive reader's *in medias res* reaction to what has been written. I think it made the student feel important to have a professor express so much concern over his efforts.

That the running commentary made during the initial reading of a paper was signally troublesome almost a fourth of the teachers testified to. After about six weeks of trying the experimental recording of comments, one instructor complained,

> I cannot read the papers, make comments as I go, and finish with any idea of a grade for the paper. This necessitates my reading the paper at least twice--once with pen in hand and once for overall assessment. If I read the paper first for assessment, I find it almost impossible not to mark errors as I note them. Once I have the pen in my hand, I make more and more marks until I end up having approached the paper in the normal fashion. I am slowly learning to restrain myself. If I allow myself to record comments on the first reading, I lose perspective the second time around--the already noted errors seem to loom much larger than they ever have in the past.

Dictating comments required periods varying from one to nine weeks to adjust to. The best technique was not a single one but, rather, an individual affair good for oneself and, perhaps, for others like him.

Undoubtedly, dictating comments while reading a paper sufficiently disrupts the average English teacher's habits to become of real concern to him. Even the smooth surface of the uncut plastic disk may be a reproach. Although at the end of the experiment most teachers said they had no compulsion to fill the blank disk regardless of the worth of the comments, others (about a quarter of them) admitted that, at the beginning, the method presented a temporary problem. One teacher, for example, after a month of trial complained,

> I feel a pressure to make at least six minutes worth of comment to every student. This may, of course, be a concomitant of my first acquaintance with their work--I normally dwell on every error

> at the beginning of a semester. Secondly, I have not yet learned to edit my remarks. Each error worth mentioning gets a full barrage.

The same instructor felt he needed both a psychological and a physical adjustment to the method:

> My approach to grading, so far [after a month], has to be much more formal than it used to be. I cannot grade on the run, nor can I pick up a few papers now and a few later. Partly because of the newness, I must gird my grading loins much more fully than in the past. And the sheer physical adjustment of all the paraphernalia--with the arrangement of my desk--makes my approach more premeditated than usual. It seems impossible to turn to a paper or two as a respite from other duties. This may have advantages in a unity of approach, but it does lead to a new kind of fatigue and it intensifies the usual period of foreboding that precedes the grading of any set of papers. Once I begin, there is no flight.

After the newness of the experiment wore off, the feeling that there was no escape from entanglement with the dictating machine was probably compounded by the knowledge that the job of reading a set of papers would now take about twice as long. Even the knowledge that, as a teacher, one was doing a more conscientious task probably could never compensate for the feeling of dread. The job had to be reckoned with. Neither a set of papers nor a modern dictating machine could ever become a coy mistress to make the "sun stand still." Thus the teacher who had established a forty-year record for promptly reading freshman papers confessed, "My greatest difficulty has been in getting themes graded promptly." And another veteran of at least twenty years of serious concern with freshman English was slightly dismayed at how dictating comments slowed her down. When at the end of a month she finally had worked out her procedure, however, she announced, "But I have already developed speed and can now grade a set of papers with relative ease and facility." Still another doubted whether the extra time would finally benefit the student, "At the present [about six weeks after beginning the experiment] I fear that the improvement made in student writing will be incommensurate with the time spent by the instructor." Regardless of the time required by the method, one teacher voiced an important psychological salve that the recorded comments offered him:

> It must help the students in the long run: the instructor is able to say exactly what he wants to say to each student on each paper. If for no other reason, the method doesn't leave the instructor frustrated after grading a set of papers. He knows he has told each student exactly what he feels

> about a paper--it's a relief.

One teacher thought that perhaps a superior method for using the recorded comment was to dispense entirely with written comment:

> I am of the opinion that we are doing too much for them /The students/; they would profit, I believe, by having to hunt out the errors that we tell them exist, thus learning what they seem incapable of now, to proofread. I believe that the recorder used alone, without written comment, would prove to be an effective and time-saving device for composition courses.

Another teacher described how he gave up the running commentary in favor of terminal remarks:

> After correcting the paper completely, I recorded a comment on the disk. Referring to pages by number, I ran through recurrent errors and sentence patterns, using the paper itself to make my points. I also gave a terminal comment on organization and content with some advice on studying. Thus, I have been trying to make the disk serve as (1) a conference with the student, and (2) a terminal comment on the paper. Depending upon the nature of the student's errors, I stress one purpose or the other.
>
> Sometimes I find that my thoughts about a paper have taken articulate form as I complete my correcting it, and I can cut the disk with few pauses; however, this is more the exception than the rule. It is extremely difficult to comment at any length to a student whose sentences are so poor that not much more can be said to him than, "Study the following thirty sections of the handbook." I have decided that the machines are not a time-saving device, however valuable they may be pedagogically. I can honestly report that, with about a fourth of the papers, I have been pleased to use the disk, to be able to discuss with the student problems that might take a half hour or more to discuss in written comments.

Recorded Comments for Advanced English Classes.

Although the experiment was carried on with freshman English classes, two teachers used recorded comments also in other classes--and were pleased with the results. Another teacher saw the possibility of such use and offered the service to two sophomore literature classes, who did not accept the invitation:

> Although I have not used the recorded comments on papers in my sophomore literature classes, I have the feeling that this approach would be perhaps even more valuable beyond the freshman level. In upper-level classes, I also like to make extended comments on organization and content. To do so in writing, however, is almost an impossibility. But to do so on record is quite within the realm of possibility. In all, I am pleased with the results and enthusiastic about the prospects.

Another instructor reinforced the use of recorded comments for upper-division papers:

> Although I strongly favor the use of disks for grading freshman papers (in fact, I felt as though I were cheating my students in the control section because I couldn't use the disks there), I think the big advantage of the disks may, finally, lie in the long paper. For example, I used two records last week to criticize a hundred pages of a master's thesis in its early-draft stages. I used about fifteen minutes of recording time in a running commentary on the first forty pages. By that time I could see the major faults of organization that the writer had fallen into. After glancing through the last sixty pages to test my hypothesis, I cut another five minutes of criticism explaining how the organization probably ought to be handled to clarify the thesis for the writer as well as the reader. I sandwiched in this reading and criticism over a period of three or four days. If I had written my comments in longhand--or even if I had typed them--I suppose I would have lengthened the job by about three hours. The distinct advantage the disk had over what would have had to be an extended personal conference was that I could approximate the effect of the personal conference in broken periods of time convenient to me, such as midnight, for example.

In spite of some of the difficulties and dissatisfactions, both teachers and students were, in general, favorably inclined toward recorded commentaries on themes. Whether better themes accompanied the liking is the business of the next chapter.

Chapter IV

STATISTICAL ANALYSIS OF RESULTS OF EXPERIMENTAL METHOD

To complement the analysis of teachers' and students' attitudes toward recorded comments reported in Chapter III, several statistical studies may give a more objective view of the effectiveness of the experiment.

The Halo of Subjectivity in Theme Evaluation.

The attempt to evaluate themes objectively is a noble ideal but a practical impossibility. English teachers, I suppose, long ago accepted this maxim even as they strove to deny it. Even though our teachers devoted their sessions of preparation for the experiment to group reading of practice themes, comparison of grades, discussion of standards, or defining categories of criteria--even though serious attention was aimed at minimizing subjectivity in grading, subjective evaluation could not be conquered. And some of the uncertainty of the experimental results must surely stem from the subjective ingredient in theme grading. Appendix G-1, "Correlation Coefficients of Judges. . .," tabulates data that substantiate the inability of our theme-graders to eliminate the subjective.

What the coefficient of correlation between two judges grading the same themes should be, I have no idea. To expect anything near perfection--that is, a correlation coefficient of 1.0--would be naive. Appendix G-1, for example, reveals that the highest coefficient of correlation for a pair of judges on the same panel scoring a paper on four categories (content, organization, sentences, mechanics) was .73--the coefficient for Judges 51 and 53; the lowest was .26 for Judges 41 and 42.

When the four categories were grouped into two, content and organization as opposed to sentences and mechanics--a similar disparity appeared. For content and organization the highest coefficient of correlation was .62 for Judges 21 and 22; the lowest was on the same panel (.03), Judges 22 and 23. For sentences and mechanics, the highest coefficient was .73 for Judges 51 and 53; the lowest was .25, for Judges 41 and 42.

When judges merely tabulated errors in sentences and mechanics without trying to weigh the significance of those errors, the coefficients of correlation showed less range and were higher: the highest was .84 for Judges 21 and 22, and the lowest was .57 for Judges 31 and 32.

In other words, two judges were more likely to agree on how many errors in sentences and mechanics a paper had than they were on how serious those errors were. But two judges were more likely to agree on the seriousness of errors in sentences and mechanics than they were on the merits of content and organization. That these two generalizations are valid is suggested by comparing the mean correlations of the three judges on each panel as they were derived from Appendix G-1:

TABLE V

Mean Correlations of
Judges Paired Within Panels

Panel	CO	SM	Err	COSM
1	.50	.61	.70	.59
2	.29	.64	.79	.46
3	.41	.52	.65	.47
4	.34	.41	.70	.37
5	.49	.58	.70	.63
Mean	.41	.55	.71	.50

On every panel the coefficient of correlation for sentences and mechanics (SM) was higher than that for content and organization (CO). Similarly the correlation between judges on actual error-count (Err) was higher than that on sentences and mechanics (SM). Even the correlation on error-count varied, of course; no one would expect it to be a perfect 1.0. For example, inspecting the count of spelling errors alone for sample individual themes showed that seldom did judges find exactly the same number of misspelled words.

Regardless of the subjective disparities, I assume that no subjective bias accumulated in any one direction in any panel: (1) because both pre-themes and post-themes were read at the same time immediately after the experiment, (2) because both pre-themes and post-themes were haphazardly intermingled without identification and distributed to the judges in packets containing about the same number of themes, (3) because no themes were identified as to author, (4) because no judge (except the one acting as research assistant) knew who else was on his panel, (5) because no judge knew how any other judge graded any paper, (6) because theme topics were the same for both pre- and post-themes, except that the topics were distributed to different students for the post-themes, (7) because no students wrote a post-theme on the same topic as their pre-themes, (8) because judges were assigned to panels according to a formula of one judge from the group of average or normal graders, balanced by one judge from the group of "hard" graders and one from the group of "easy" graders.

Comparison of Pre-STEP and Post-STEP Results.

The most objective measurement of the effectiveness of recorded comments--but not necessarily the best measurement--was the results of comparing performance on Pre-STEP[1] and Post-STEP[2] examinations, which are summarized in Appendix G-2.

All students were divided into five groups corresponding to the five panels of judges who read their themes. For both experimental and control groups, the mean Pre-STEP score for each panel was compared with the mean Post-STEP score. Then the percentage of increase (or decrease) for each panel was computed. Finally the mean percentage of increase or decrease for all five panels was figured to yield an over-all, gross measure.

The percentage of increase for all sections was almost identical for experimental and control groups: rounded off, the figure was 1.7% increase for both. According to this measure, then, both the experimental and the control groups showed the same amount of progress. Recorded comments, in general, had no effect one way or another.

When only students in Regular Sections, made up of the bottom ninety percent, were studied--approximately the same results were apparent. Except that Regular Sections alone showed slightly more improvement than all sections combined, the Experimental Regular Sections were not significantly superior to the Control Regular Sections. Experimental Regular Sections showed a mean percentage improvement of 2.2; the Control Regular Sections, an improvement of 2.0%.

The upper ten percent of the students, in Superior Sections, did not entirely follow the pattern: Experimental Superior Sections improved only .5%; but Control Superior Sections improved 1.1%. Although the figure for the control group is twice that for the experimental and therefore seems relatively significant, both figures are probably too small to draw any conclusions.

---

[1] Sequential Tests of Educational Progress, Writing Form 1-A.

[2] Sequential Tests of Educational Progress, Writing Form 1-B.

Pre-Theme and Post-Theme Results: COSM.

Figures for computing percentage of improvement in theme writing are presented in Appendix H, from which Tables VI, VII, and VIII are derived. Data were gathered for two populations: (1) The larger population, labelled "all students," includes every student enrolled in experimental and control sections for whom a complete set of results was available; in this all-student population the number enrolled in an experimental section might be larger or smaller than the number enrolled in the matching control section. (2) The smaller population, labelled "paired students," includes only those students in experimental sections who could be coupled with a counterpart in the matching control section on the basis of Pre-STEP scores; the paired-student population, therefore, was selective, and the number in the experimental group was identical to the number in the control group.

Because the paired-student population is more likely to assure that both the experimental group and the control group started the experiment on the same level of achievement, I shall rely on that population to educe conclusions. But data for the all-student population are also presented for those who do not agree with such an an assumption.

As Table VI reveals, the percentage of improvement based on over-all theme scores (content, organization, sentences, and mechanics) was not significantly different in all sections for the experimental (+2.9%) and control (+3.3%) groups.

TABLE VI

Pre-Theme to Post-Theme
Percentage of Improvement
On Mean Scores of COSM, CO, SM
Paired Students versus All Students
Summary of Appendix H: All Sections

	Experimental Group		Control Group	
	Pairs	All Students	Pairs	All Students
COSM*	+2.9%	+2.7%	+3.3%	+2.3%
CO	+6.1%	+4.3%	+4.8%	+3.7%
SM	+0.9%	+0.2%	+0.6%	+1.2%

*C - Content; O - Organization; S - Sentences; M - Mechanics

A possibly significant difference, however, does appear when only content and organization are considered: the percentage of improvement was +6.1% for the experimental group as opposed to +4.8% for the control. In other words, throughout all sections recorded comments had a slight edge over the traditional method in improving students' content and organization.

Table VII, though, casts a different light on the validity of such a conclusion.

TABLE VII

Pre-Theme to Post-Theme
Percentage of Improvement
On Mean Scores of COSM, CO, SM
Paired Students versus All Students

Summary of Appendix H: Regular Sections

	Experimental Group		Control Group	
	Pairs	All Students	Pairs	All Students
COSM*	+2.7%	+2.5%	+3.6%	+2.8%
CO	+4.7%	+4.6%	+6.0%	+4.3%
SM	+0.6%	+0.2%	+0.1%	+0.5%

*C - Content; O - Organization; S - Sentences; M - Mechanics

When only the students in Regular Sections are considered, the control and experimental groups exchange positions in showing improvement in the categories of content and organization: the control group showed +6.0% improvement as compared to the experimental group's +4.7%. This reversal suggests that the edge recorded comments had over the traditional method in improving content and organization throughout all sections was not provided by the Regular Sections.

That the Superior Sections provided this edge Table VIII testifies to.

TABLE VIII

Pre-Theme to Post-Theme
Percentage of Improvement
On Mean Scores of COSM, CO, SM
Paired Students versus All Students

Summary of Appendix H: Superior Sections

	Experimental Group		Control Group	
	Pairs	All Students	Pairs	All Students
COSM*	+5.1%	+3.6%	+2.9%	+4.0%
CO	+8.2%	+6.1%	+3.5%	+2.0%
SM	+0.9%	+0.2%	+2.0%	+1.6%

*C - Content; O - Organization; S - Sentences; M - Mechanics

In the Superior Sections only, the experimental paired-student population showed +8.2% inprovement in contrast to the control all-student population's +3.5% improvement; that is, in the Superior Sections the experimental group's rate of improvement in content and organization was nearly two and one-half times greater than the control group's rate. In the all-student population, the experimental group's rate of improvement was more than three times that of the control group's rate.

If any conclusion can be drawn from this evidence, it is probably this: recorded comments have recognizable value among superior students in improving their content and organization.

That conclusion partially bears out what sixty percent of the teachers and fifty-six percent of the students reported about the helpfulness of recorded comments in improving content and organization more than sentences and mechanics (see Chapter III and Appendix F-2a). Both students' and teachers' attitudes, however, must be modified now to refer only to superior students, not to all of them. Probably, however, this hypothesis requires further research.

Pre-Theme and Post-Theme Results: Error Count.

Tables VI, VII, and VIII show that, in the paired-student population, the experimental group in all sections and Regular Sections had a slightly higher percentage of improvement in sentences and mechanics than the control group did. In the Superior Sections, however, the control group had a higher percentage of improvement (experimental--+0.9%; control--+2.0%). The values, in any case, are so small as to be insignificant, one way or the other.

The scores on sentences and mechanics, on which Tables VI, VII, and VIII are based, represented teachers' judgments about errors made in ten categories: sentence sense, agreement, reference, verbs, case, modifiers, parallelism, spelling, capitalization, and punctuation. As they read the papers, the teachers tallied the errors they discovered in each of the categories. The first seven categories were regarded as errors in sentences and the last three as errors in mechanics. Scores students received in sentences and mechanics were not based on gross error count only, but were assessed according to a rough scale based on about a four-hundred-word theme and according to a rough scale of seriousness (errors in sentence sense, for example, were judged more serious than errors in case, etc.).

When as much teachers' judgment as possible was eliminated, however, and evidence of gross count of errors is examined--a different pattern evolves: one that bolsters the evidence that recorded comments have special value for superior students in improving their content and organization. Of the ten categories under which errors were tabulated, seven--namely, sentence sense, agreement, reference, verbs, case, modifiers, and parallelism--seem more closely related to content and organization than the other three categories do. In other words, those seven categories are more likely to require the same qualities of mind for improvement as those required in content and organization.

Table IX (summarized from Appendix J) merely hints at the possible relationship of content-organization and the seven categories of errors.

TABLE IX

Pre-Theme to Post-Theme
Percentage of Improvement
On Mean Error-Counts: Ten Categories*
Paired Students versus All Students

Summary of Appendix J

	Experimental Group		Control Group	
	Pairs	All Students	Pairs	All Students
All Sections	-4.8%	-5.2%	-0.4%	+0.6%
Regular Secs.	-7.1%	-5.1%	+0.4%	+1.8%
Superior Secs.	-10.3%	-18.2%	-14.4%	-13.8%

*Ten Categories: Sentence Sense, Agreement, Reference, Verbs, Case, Modifiers, Parallelism, Spelling, Capitalization, Punctuation

In all ten categories of errors, only the experimental group in the Superior Sections showed a higher percentage of improvement than the control group. Because we are concerned in this study only with relative standing of experimental and control groups, the fact that the percentage of improvement is in negative figures and therefore implies "deterioration" is irrelevant. What looks like deterioration may in fact be progress masked under higher gross errors because the post-themes contained more words and so presented more opportunities for making errors.

Table X (derived from Appendix J) refines the hint by revealing that the experimental group in the Superior Sections (compared with the control group) had a little more than two times the percentage of improvement in the seven categories. The experimental group in the Superior Sections, it will be remembered, held about the same advantage over the control group when improvement in content and organization was under examination.

TABLE X

Pre-Theme to Post-Theme
Percentage of Improvement
On Mean Error-Counts: Seven Categories*
Paired Students versus All Students

Summary of Appendix J

	Experimental Group		Control Group	
	Pairs	All Students	Pairs	All Students
All Sections	-2.7%	-0.6%	+7.8%	+5.6%
Regular Secs.	-2.9%	+0.7%	+9.5%	+8.2%
Superior Secs.	-18.5%	-11.0%	-38.2%	-14.0%

*Seven Categories: Sentence Sense, Agreement, Reference, Verbs, Case, Modifiers, Parallelism

Table XI (condensed from Appendix J) merely confirms the conclusion. For three categories of errors related almost wholly to mechanics and hardly at all to content and organization, the experimental group was consistently inferior to the control group.

TABLE XI

Pre-Theme to Post-Theme
Percentage of Improvement
On Mean Error-Counts: Three Categories*
Paired Students versus All Students

Summary of Appendix J

	Experimental Group		Control Group	
	Pairs	All Students	Pairs	All Students
All Sections	-7.7%	-8.6%	+0.3%	-2.8%
Regular Secs.	-9.1%	-8.4%	-1.2%	-3.3%
Superior Secs.	-14.1%	-25.2%	-7.0%	-12.6%

*Three Categories: Spelling, Capitalization, Punctuation

Thus Table XI strongly suggests that for all students--but especially for the average and below-average--the traditional method of marginal notes is superior to recorded comments to improve students' spelling, capitalization, and punctuation. Recorded comments are best reserved to explain complex ways of strengthening organization, content, and sentence structure. Marginal notes are sufficient to point out non-complex errors in usage that rapid reference to a handbook can clarify.

Chapter V

CONCLUSION

The questions listed at the beginning of Chapter I can now receive at least partial answers--and tentative ones.

The Basic Question.

In Chapter I, the central question that the study sought an answer to was phrased as follows:

Will teacher commentary, recorded on a plastic disk at 33 1/3 rpm and made available for replay to students for revision, improve their subsequent, original writing done within twelve weeks?

For the average student the answer is "yes"; but the answer must be qualified: recorded commentary, for the average student, can improve his written composition over a twelve-week period; but the improvement is likely to be no more dramatic than that caused by the traditional method of marginal notation or final short, written comments.

Recorded Commentary: Its Effect on Content and Organization.

Two of the "related, but subsidiary questions" listed in Chapter I were:

What effect does recorded commentary have on students' content and organization as opposed to their sentence structure and mechanics?

What is the relative effectiveness of the experimental method on above-average students' writing as opposed to average and below-average students' writing?

The experiment provided a single, tentative answer to both questions. The single, important improvement that recorded commentary can bring about is in better content and organization, and that improvement seems to be restricted to superior students only.

Where the traditional method brought about 2%-3.5% improvement in content and organization among superior students in the control group, recorded commentary brought about 6.1%-8.2% among the superior students in the experimental group. In other words the experimental group had improvement that was between two and three times that of the control group.

Recorded Commentary: Its Effect on STEP Writing Performance

In Chapter I, the question was, What effect does recorded commentary have on students' performance on a standardized, objective test of writing ability?

The answer is short: none. The amount of improvement over the twelve weeks was unaffected by whether students received recorded commentary or only marginal notes of the traditional method. Whether students were in Superior Sections or in Regular Sections also made no significant difference.

Recorded Commentary: Its Effect on Selected Errors.

One pre-experiment question was, What effect does recorded commentary have on selected errors that students make? The evidence for a clear-cut answer is contradictory.

Teachers' judgments about the seriousness of errors in sentences and mechanics show no difference between the experimental and control groups.

Teachers' count of errors seems to be almost wholly unreliable because the free writing situation is so full of variables as to make mere count unimportant.

Basic Recommendations.

(1) I am satisfied that the advantage of recorded comments for improving the content and organization of superior students is a real advantage. On the freshman level, therefore, I recommend that freshman English teachers use recorded comments in superior sections, in those colleges where superior, or honors, freshmen are segregated in special classes. Although my recommendation leans on an administrative device rather than an academic desirability, it avoids selecting only the best students from a heterogeneous section for preferential treatment.

(2) I further suggest that the recorded commentary will be especially useful in upper division classes and graduate classes to criticize term papers, theses, dissertations, and even essay examinations. Although this study did not develop any evidence for such use, I think it is a logical extension of my first recommendation.

Secondary Recommendations.

In Chapter IV I talked briefly about the "halo of subjectivity" surrounding the evaluation of students' themes by teachers. I think this whole problem needs study, but the one area I wish to recommend for further study would be important in helping teachers work toward standardization. These are some of the questions that I think need answering:

How high can one expect the coefficient of correlation to be for two judges grading the same set of themes? Or, conversely, how low a coefficient of correlation should be regarded as unsatisfactory? The object, I should think, would be to try to discover a norm derived from comparing teachers from a cross-section of types of colleges from various parts of the nation.

Answers to the preceding questions, in one sense, may be irrelevant to the grading within a particular college or university. What is needed is some measure of variability of judgment within a group of teachers. In other words, what should be the limits of variability of judgment within a given group of teachers? That is, when can one say to an individual teacher, your judgment is unacceptably harsh (or light, uncertain, etc.) when compared to others in our group of instructors? Related to these is: when can a supervisor know that practice-grading as an in-service training project has continued long enough--or must be continued longer, or must be resumed, etc.?

SELECTED REFERENCES

Clark, W. A., "Neither a Margin Scribbler Nor a Juggler of Numbers," *English Journal*, College Edition (February 1939), 133-138.

Cohen, N., "Correcting Compositions Without a Pencil," *English Journal* (December 1950), 579-580.

Ford, P. M., "Lay Readers in the High School Composition Program: Some Statistics," *English Journal* (November 1961), 522-528.

Kilburn, P. E., "Every Man His Own Pedagogue: A Project in the Teaching of Freshman Composition," *The Journal of Higher Education* (February 1962), 89-95.

Laird, Charlton, "College English During the Flood," *College English* (December 1956), 131-138.

Lumsden, R., "Evanston, Illinois, Township High School Adds to Its Program: The Use of Dictation Machines in Grading English Themes," *The Bulletin of the National Association of Secondary School Principals* (January 1962), 223-226.

Lyman, R. L., *Summary of Investigations Relating to Grammar, Language, and Composition* (Chicago: University of Chicago Press, 1929).

Reeves, Bruce, "The Dictation Machine as an Aid in Correcting Compositions," Speech at Annual Convention of National Council of Teachers of English, San Francisco, 23 November 1963.

APPENDIX A-1

REGULAR SECTIONS

English 131. Composition and Rhetoric: Exposition

Outline of Study

Basic Textbooks:

Anderson, Wallace L., and Stageberg, Norman C., Introductory Readings on Language (New York: Holt, Rinehart, and Winston, Inc., 1962)

Hamilton, Edith, Mythology (New York: The New American Library of World Literature, Inc., copyright 1942 by Edith Hamilton)

Kallsen, T. J., and McCoy, D. E., Rhetoric and Reading: Order and Idea (New York: Dodd, Mead & Company, 1963)

Leggett, Glenn; Mead, C. David; and Charvat, William, Prentice-Hall Handbook for Writers (Third Ed., Edgewood Cliffs, N. J.: Prentice-Hall, Inc., 1960)

Webster's Seventh New Collegiate Dictionary (Springfield, Mass.: G. C. Merriam Company, 1963)

First Half-Semester

Part One: Three Weeks

Rhetorical Emphasis: (a) Purpose and Process in Communication, and (b) General Guidelines to the Writing of College Compositions

Linguistic Emphasis: The Nature of Language

Composition: (a) First Paper, In-Class Theme, Diagnostic First Week; (b) Second Paper, Out-of-Class Theme, Second Week

Spelling Test I: Second Week

Part Two: Five Weeks

Rhetorical Emphasis: (a) Effect of Analysis on the Language and Meaning, (b) Principles of Analysis and Definition, (c) Analysis and Organization of Composition

Linguistic Emphasis: (a) Forms and Meanings of Words, (b) Metaphorical Language, (c) Semantics

Composition: (a) Third Paper, In-Class Theme, Fourth Week; (b) Fourth Paper, Out-of-Class Theme, Sixth Week; (c) Fifth Paper, In-Class Theme, Eighth Week, Extended Definition as Midsemester Test

Spelling Test II: Fifth Week

Last Half-Semester

Part Three: Eight Weeks

Rhetorical Emphasis: Basic Ways to Divide and Classify

Linguistic Emphasis: (a) The Oral Language, Speech Sounds, and Geography, (b) Writing as Process, (c) Reading as Process

Composition: (a) Sixth Paper, Out-of-Class Theme, Ninth Week; (b) Seventh Paper, Out-of-Class Theme, Tenth Week; (c) Eighth Paper, In-Class Theme, Twelfth Week; (d) Ninth Paper, Out-of-Class Theme, Fourteenth Week; (e) Tenth Paper, In-Class Theme, Sixteenth Week,

Final Examination on Hamilton's Mythology

English 132. Composition and Rhetoric: Argument

Outline of Study

Basic Textbooks:

Anderson and Stageberg, *op. cit.*

Kallsen and McCoy, *op. cit.*

Leggett, Mead, and Charvat, *op. cit.*

Rouse, W. H. D. (trans.) Homer's *The Iliad* (New York: The New American Library of World Literature, Inc., copyright by Thomas Nelson and Sons, Ltd., 1938)

Selected Novel

Webster's Seventh New Collegiate Dictionary, *op. cit.*

First Half-Semester

Part One: Four Weeks

Rhetorical Emphasis: (a) Emotional Proof, (b) Logical Proof, (c) The Research Paper

Linguistic Emphasis: (a) Relations of Clear Thinking to Language, (b) Acceptable Usage

Composition: (a) First Paper, In-Class Theme, Argument by Reasons, First Week; (b) Second Paper, Out-of-Class Theme, Argument by Deduction, Third Week; (c) Third Paper, In-Class Theme, Argument by Induction, Fourth Week

Part Two: Four Weeks

Rhetorical Emphasis: (a) Evaluation--Patterns and Standards, (b) The Research Paper (continued)

Linguistic Emphasis: (a) Standards of Pronunciation and Usage, (b) Reading and Education

Composition: (a) Fourth Paper, Out-of-Class Theme, Evaluation, Sixth Week; (b) Fifth Paper, In-Class Theme, Evaluative Paper on The Iliad as Midsemester Examination, Eighth Week

Last Half-Semester

Part Three: Five Weeks

Rhetorical Emphasis: (a) Argument of Policy, (b) The Research Paper (continued)

Linguistic Emphasis: (a) Dictionaries and Usage, (b) Speed and Comprehension in Reading, (c) Style in Writing, (d) English as a World Language

Composition: (a) Sixth Paper, Out-of-Class Theme, Argument Problem and Solution, Tenth Week; (b) Seventh Paper, In-Class Theme, Argument by Presenting General Issues, Eleventh Week; (c) Eighth Paper, Out-of-Class Theme, Argumentative or Evaluative Research Paper, Thirteenth Week

Part Four: Three Weeks

Rhetorical Emphasis: Aesthetic Discourse as Exemplified in a Novel

Linguistic Emphasis: Suited to the Chosen Novel

Composition: Tenth Paper, In-Class Theme, Evaluative or Argumentative Theme on Assigned Novel, Sixteenth Week

Usage Examination: Examination Week

NOTE: For the project, the first in-class theme (pre-theme) became an argumentative paper written from a choice of propositions furnished the student (see Appendix B for sets of topics). The last theme, an in-class paper (post-theme), also argumentative, with the same set of propositions rearranged (see Appendix B).

APPENDIX A-2

SUPERIOR SECTIONS

English 133: *Composition and Rhetoric: Exposition and Argument*

Outline of Study

Basic Textbooks:

Anderson and Stageberg, *op. cit*.

Kallsen and McCoy, *op. cit*.

Leggett, Mead, and Charvat, *op. cit*.

Rouse, *op. cit*.

Selected Novel

Webster's Seventh New Collegiate Dictionary, *op. cit.*

First Half-Semester

Part One: One and One-half Weeks

Rhetorical Emphasis: See Part One of Outline for English 131

2

Linguistic Emphasis: See Part One of Outline for English 131

Composition: First Paper, In-Class Theme, Choice of First or Second Paper Given in Part One of Outline for English 131

Part Two: Two and one-half Weeks

Rhetorical Emphasis: See Part Two of Outline for English 131

Linguistic Emphasis: See Part Two of Outline for English 131

Composition: (a) Second Paper, Out-of-Class Theme, Classification, Third Week; (b) Third Paper, In-Class Theme, Division or Extended Definition, Fourth Week

Part Three: Four Weeks

Rhetorical Emphasis: See Part Three of Outline for English 131

Linguistic Emphasis: See Part Three of Outline for English 131

Composition: (a) Fourth Paper, Out-of-Class Theme, Causal Sequence or Comparison, Seventh Week; (b) Fifth Paper, In-Class Theme on The Iliad with Expository Organization, Eighth Week

Second Half-Semester

Part Four: One and one-half Weeks

Rhetorical Emphasis: See Part One of Outline for English 132

Linguistic Emphasis: See Part One of Outline for English 132

Composition: Sixth Paper, Out-of-Class Theme, Argument by Deduction, Ninth Week

Part Five: One and one-half Weeks

Rhetorical Emphasis: See Part Two of Outline for English 132

Linguistic Emphasis: See Part Two of Outline for English 132

Composition: Seventh Paper, In-Class Theme, Evaluation, Eleventh Week

Part Six:

Rhetorical Emphasis: See Part Three of Outline for English 132

Linguistic Emphasis: See Part Three of Outline for English 132

Composition: (a) Eighth Paper, Out-of-Class Theme, Argument by Problem and Solution, Based on Research, Thirteenth Week; (b) Ninth Paper, Out-of-Class Theme, Argument on General Issues, Based on Research, Fourteenth Week

Part Seven: Two Weeks

Rhetorical Emphasis: See Part Four of Outline for English 132

Linguistic Emphasis: See Part Four of Outline for English 132

Composition: Tenth Paper, In-Class Theme, Evaluative or Argumentative Essay on Selected Novel, Examination Week.

English 134: Composition and Rhetoric: Literature

Outline of Study

First Half-Semester

Part One: Four and one-half Weeks

Poetry: Laurence Perrine, Sound and Sense (Second Ed., New York: Harcourt, Brace and World, 1963)

First Paper: In-Class Theme, Critical Essay on Poetry, Second Week

Second Paper: Out-of-Class Theme, Criticism or Explication of Poetry, Fourth Week

Part Two: Three Weeks

Short Stories: C. L. Cline (ed.), The Rinehart Book of Short Stories (New York: Holt, Rinehart, and Winston, Inc., RE59)

Third Paper: Out-of-Class Theme, Criticism of Short Story, Seventh Week

Fourth Paper: Midsemester Essay, In-Class Theme

Second Half-Semester

Part Three: Four Weeks

Two Shakespearean Plays: Alan S. Downer (ed.), Othello and Twelfth Night (New York: Holt, Rinehart, and Winston, Inc., RE92)

Fifth Paper: In-Class Theme, Critical Essay on One or Both Plays, Twelfth Week

Part Four: One Week

One Modern Play: Six Great Modern Plays (New York: Dell Books, 798LE)

Sixth Paper: Out-of-Class Theme (Optional with Either Part Four or Part Five), Thirteenth or Fourteenth Week

Part Five: Three Weeks

Selected Novel

Seventh Paper: Out-of-Class Theme, Research Paper Growing out of Novel Study, Fifteenth Week

Eighth Paper: Final Examination Essay, Examination Week

NOTE: In the project English 134 students wrote pre- and post-themes on topics listed in Appendix B.

APPENDIX B

Topics for Pre- and Post- Themes

The first themes assigned in English 132 and 134 for the project (designated as pre-themes) and the last themes assigned (known as post-themes) had identical directions for the student to follow:

> You have fifty minutes to write an argument either for or against one of the questions listed below. Your argument should run at least four hundred words. You need not hand in an outline, but you will be graded on organization. If you wish to correct a word, phrase, or sentence, run a line neatly through the passage you want to delete and then write legibly above the crossed-out passage. CHOOSE ONE QUESTION TO ARGUE:

Then, questions listed in Set 1 were assigned to students meeting sections during the first two class periods; Set 2 to students meeting sections during the next two class periods, etc. In this way, information about the topics could not be effectively passed on from one period to the next.

APPENDIX B-1

Pre-Theme Topics

Set 1:

1. Should the federal government establish a bureau to censor books and magazines?

2. Should the federal government subsidize higher education for the more intelligent students?

3. Should students desiring a bachelor's degree from a Texas college be required to earn two semester of credit in government?

4. Should Stephen F. Austin State College drop scholarships for athletes participating in intercollegiate athletics?

Set 2:

1. Should each state in the USA be allowed two senatorial seats in the Congress?

2. Should the United States withdraw from the United Nations?

3. Should six semester hours of literature be required for a bachelor's degree at Stephen F. Austin State College?

4. Should students living in dormitories be allowed to park their cars anywhere on campus?

Set 3:

1. Should the United States government operate all railroads, air lines, bus companies, and truck lines in interstate commerce?
2. Should the United States release control of the Panama Canal to an international commission appointed by the United Nations?
3. Should freshman English be required of all students at Stephen F. Austin State College?
4. Should all students at Stephen F. Austin State College be required to buy a student activity ticket admitting them to all college-sponsored extra-curricular activities?

Set 4:

1. Should the United States adopt uniform divorce laws throughout all the states?
2. Should the federal government declare all poll taxes illegal?
3. Should foreign language be required for every bachelor's degree awarded at Stephen F. Austin State College?
4. Should women living in dormitories at Stephen F. Austin State College be required to buy cafeteria meal tickets?

Set 5:

1. Should the United States adopt a uniform traffic code throughout all the states?
2. Should the United States establish a bureau to set up standards of language usage?
3. Should physical activity courses be required for a bachelor's degree at Stephen F. Austin State College?
4. Should men living in dormitories at Stephen F. Austin State College be required to buy cafeteria meal tickets?

Set 6:

1. Should religion be taught in state-supported colleges?
2. Should the United States government financially support a national symphony orchestra?
3. Should all students at Stephen F. Austin State College be required to study at least one course in art appreciation?
4. Should the time-card system for registration at Stephen F. Austin State College be changed?

Set 7:

1. Should atheists be allowed to teach in state-supported schools?
2. Should the United States government financially support a national ballet company?
3. Should all students at Stephen F. Austin State College be required to study at least one course in music appreciation?
4. Should commuting students at Stephen F. Austin State College be given reserved parking zones?

Set 8:

1. Should the voting age be lowered to eighteen years in the United States?
2. Should the United States discontinue Radio Free Europe?
3. Should two years of science be required for every bachelor's degree awarded at Stephen F. Austin State College?
4. Should faculty members at Stephen F. Austin State College be given specially reserved parking spaces?

Set 9:

1. Should the federal government establish a bureau to censor movies and plays?
2. Should the United States discontinue the Peace Corps?
3. Should students wanting a bachelor's degree from a Texas College be required to earn two semesters of credit in American history?
4. Should the library at Stephen F. Austin State College be kept open Sunday afternoons and evenings?

APPENDIX B-2

Post-Theme Topics

Set 1: Same as Pre-Theme, Set 5
Set 2: Same as Pre-Theme, Set 4
Set 3: Same as Pre-Theme, Set 6
Set 4: Same as Pre-Theme, Set 2
Set 5: Same as Pre-Theme, Set 1
Set 6: Same as Pre-Theme, Set 9
Set 7: Same as Pre-Theme, Set 8
Set 8: Same as Pre-Theme, Set 7
Set 9: Same as Pre-Theme, Set 3

APPENDIX B-3

Assignment of Sets to Class Periods

Set 1: 8 and 9 Monday-Wednesday-Friday classes

Set 2: 10 and 11 Monday-Wednesday-Friday classes

Set 3: 12 and 1 Monday-Wednesday-Friday classes

Set 4: 2 Monday-Wednesday-Friday classes

Set 5: 7:45 and 9:10 Tuesday-Thursday classes

Set 6: 10:35 and 12:00 Tuesday-Thursday classes

Set 7: 1:25 and 2:50 Tuesday-Thursday classes

Set 8: 4:10 and 5:30 Tuesday-Thursday classes

Set 9: Make-up

APPENDIX C-1

Stephen F. Austin State College

ERROR-COUNT SHEET

____________________ ____________________

student's code number instructor's name

1. SENTENCE SENSE
   fragments, comma splices, fused sentences 1. ____
2. AGREEMENT
   subject-verb, pronoun-antecedent 2. ____
3. REFERENCE
   vagueness, ambiguity, indefiniteness 3. ____
4. VERBS
   tense, principal parts 4. ____
5. CASE
   nominative, objective, possessive 5. ____
6. MODIFIERS
   comparative-superlative forms, danglers 6. ____
7. PARALLELISM
   illogical construction in lists, comparisons 7. ____
8. SPELLING
   unconventional variants 8. ____
9. CAPITALIZATION
   unconventional particular cases 9. ____
10. PUNCTUATION
    unconventional formal punctuation 10. ____

NOTE: Unconventional formal punctuation-careless omission of periods at ends of sentences and to mark abbreviations; omission of commas in dates, addresses, series (optional before and), nonrestrictives, mild exclamations; hyphens in compounding; quotation marks for direct quotations, titles; etc.

Directions:

1. DO NOT put any marks on the theme itself.
2. On the error count sheet write (a) the student's code number in place of his name, (b) your last name in the blank calling for your name, and (c) the actual number of errors made in the ten categories listed. For example, if the theme had three misspellings, you should enter the figure 3 in Blank 8. The figures to be entered on the error count sheet are not percentages; they do not represent grades; they are an actual count of types of errors. The ten categories do not exhaust the types of errors you are likely to find on a paper; they are only selected errors.

APPENDIX C-2

CRITERIA-AND-SCORE SHEET FOR RATING THEMES

____________________ CON ___
student's code number ORG ___ ___
SEN ___
____________________ MEC ___
instructor's name TOTAL ___

C CONTENT 30
Examples and illustrations are RELEVANT to the PURPOSE. Main ideas are developed PROPORTIONATELY, with none left as unsubstantiated generalizations or unauthenticated conclusions. Material is sufficiently SPECIFIC and CONCRETE to lend authority to writing. Paper fulfills ASSIGNMENT. ___

O ORGANIZATION 30
INTRODUCTION gains attention, states central idea, shows mode of development. DISCUSSION has main divisions logically ordered. Paragraphs are held together with effective SUMMARY of main ideas and REFERENCE to purpose so that expository, evaluative, or argumentative nature of paper is emphasized. ___ ___

S SENTENCES 20
VARIETY is achieved without ineffective FRAGMENTS, COMMA SPLICES, or FUSED sentences. CLARITY is promoted without MISPLACED or DANGLING modifiers, SPLIT or NONPARALLEL construction, ambiguous REFERENCE, SHIFTS in viewpoint or tense, VERBIAGE, OVER-ELEGANCE, MIXED FIGURES. EFFECTIVENESS is increased by use of ACTIVE VERBS, IDIOMATIC EXPRESSION, ORIGINALITY and ECONOMY of diction. APPROPRIATENESS in usage is apparent in PROPER TENSES and PRINCIPAL PARTS in subject-verb and pronoun-antecedent AGREEMENT, and in standard inflections. ___

M MECHANICS 20
PUNCTUATION aids communication, clarifies thought, adheres to conventions of style book, does not draw reader's attention to its vagaries. SPELLING and CAPITALIZATION is standard. ___ ___

	C	O	S	M	C-O	S-M	C-O-S-M
A	27-30	27-30	18-20	18-20	54-60	36-40	90-99
B	24-26	24-26	16-17	16-17	48-53	32-35	80-89
C	21-23	21-23	14-15	14-15	42-47	28-31	70-79
D	18-20	18-20	12-13	12-13	36-41	24-27	60-69
F	0-17	0-17	0-11	0-11	0-35	0-23	0-59

APPENDIX D-1

DETAIL OF

TEACHERS' SCORES AND DEVIATIONS

ON ELEVEN PRACTICE THEMES

	Theme 1		Theme 2		Theme 3		Theme 4	
Tchr	Score	Dev.	Score	Dev.	Score	Dev.	Score	Dev.
A	42	+ 2.7	74	- 6.1	67	-11.6	43	- 8.3
B	35	- 4.3	68	+ 0.1	78	- 0.6	61	+ 9.7
C	40	+ 0.7	75	+ 7.1			63	+11.7
D	25	-14.3	56	-11.9	80	+ 1.4	37	-14.3
E	59	+19.7	71	+ 3.1	90	+11.4	45	- 6.3
F	52	+12.7	70	+ 2.1	86	+ 7.4	54	+ 2.7
G	48	+ 8.7	63	- 4.9	83	+ 4.4	53	+ 1.7
H	36	- 3.3	64	- 3.9	75	- 3.6	60	+ 8.7
I	41	+ 1.3	69	+ 1.1	94	+15.4	42	- 9.3
J	12	-27.3	63	- 4.9	88	+ 9.4	70	+18.7
K	46	+ 6.7	69	+ 1.1	69	- 9.6	69	+17.7
L	34	- 5.3	72	+ 4.1	69	- 9.6	63	+11.7
M	40	+ 0.7	76	+ 8.1	72	- 6.6	70	+18.7
N	48	+ 8.7	73	+ 5.1	82	+ 3.4	49	- 2.3
O	31	- 8.3	55	-12.9	68	-10.6	60	+ 8.7
Mean	39.3		67.8		78.6		51.3	
Average deviation	8.3		4.9		7.5		9.4	

(cont. on next page)

DETAIL OF
TEACHERS' SCORES AND DEVIATIONS
(continued from preceding page)

	Theme 5		Theme 6		Theme 7		Theme 8	
Tchr	Score	Dev.	Score	Dev.	Score	Dev.	Score	Dev.
A	29	- 7.6	80	+ 3.7	41	-14.6	77	- 5.7
B	34	- 2.6	68	- 8.3	56	+ 0.4	86	+ 3.3
C	38	+ 1.4	68	- 8.3	55	- 0.6	86	+ 3.3
D	14	-22.6	80	+ 3.7	61	+ 5.4	63	-19.7
E	38	+ 1.4	82	+ 5.7	67	+11.4	92	+ 9.3
F	48	+11.4	81	+ 4.7	85	+19.4	93	+10.3
G	37	+ 0.4	90	+13.7	66	+10.4	88	+ 5.3
H	33	- 3.6	67	- 9.3	55	- 0.6	84	- 1.3
I	46	+ 9.4	71	- 5.3	46	- 9.6	73	- 9.7
J	30	- 6.6	85	+ 8.7	70	+14.4	86	+ 3.3
K	45	+ 8.4	69	- 7.7	50	- 5.6	84	- 1.3
L	28	- 8.6	86	+ 9.7	39	-16.6	78	- 4.7
M	47	+10.4	65	-11.3	46	- 9.6	80	- 2.7
N	44	+ 7.4	84	+ 7.7	40	-15.6	86	+ 3.3
O	38	+ 1.4	69	- 7.3	57	+ 1.4	85	+ 2.3
Mean	36.6		76.3		55.6		82.7	
Average Deviation	6.9		7.7		9.0		5.7	

(cont. on next page)

DETAIL OF
TEACHERS' SCORES AND DEVIATIONS
(Continued from preceding page)

	Theme 9		Theme 10		Theme 11	
Tchr	Score	Dev.	Score	Dev.	Score	Dev.
A	51	- 9.7	80	+ 1.9	44	-11.1
B	66	+ 5.3	87	+ 8.9	64	+ 8.9
C	60	- 0.7	75	- 3.1	63	+ 7.9
D	49	-11.7	76	- 2.1	46	- 9.1
E	54	- 6.7	83	+ 4.9	60	+ 4.9
F	62	+ 1.3	74	- 4.1	68	+12.9
G	63	+ 2.3	76	- 2.1	58	+ 2.9
H	64	+ 3.3	81	+ 2.9	71	+15.9
I	56	- 4.7	74	- 4.1	47	- 8.1
J	79	+18.3	88	+ 9.9	43	-12.1
K	63	+ 2.3	75	- 3.1	56	+ 0.9
L	59	- 1.7	83	+ 4.9	47	- 8.1
M	69	+ 8.3	75	- 3.1	56	+ 0.9
N	58	- 2.7	58	-20.1	51	- 4.1
O	57	- 3.7	87	+ 8.9	53	- 2.1
Mean	60.7		78.1		55.1	
Average Deviation	5.5		5.6		7.3	

APPENDIX D-2

SUMMARY OF

TEACHERS' DEVIATIONS FROM MEAN SCORES ON ELEVEN PRACTICE THEMES

Themes

	1	2	3	4	5	6	7	8	9	10	11
Mean Score[1]	39	68	79	51	37	76	56	83	61	78	55
Average Deviation[2]	8	5	8	9	7	8	9	6	5	6	7

Themes

Tchr	1	2	3	4	5	6	7	8	9	10	11
A	+ 3	+ 6	-12	- 8	- 8	+ 4	-15	- 6	-10	+ 2	-11
B	- 4	0	- 1	+10	- 3	- 8	0	+ 3	+ 5	+ 9	+ 9
C	+ 1	+ 7		+12	+ 1	- 8	- 1	+ 3	- 1	- 3	+ 8
D	-14	-12	+ 1	-14	-23	+ 4	+ 5	-20	-12	- 2	- 9
E	+20	+ 3	+11	- 6	+ 1	+ 6	+11	+ 9	- 7	+ 5	+ 5
F	+13	+ 2	+ 7	+ 3	+11	+ 5	+19	+10	+ 1	- 4	+13
G	+ 9	- 5	+ 4	+ 2	0	+14	+10	+ 5	+ 2	- 2	+ 3
H	- 3	- 4	- 4	+ 9	- 4	- 9	- 1	+ 1	+ 3	+ 3	+16
I	+ 1	+ 1	+15	- 9	+ 9	- 5	-10	-10	- 5	- 4	- 8
J	-27	- 5	+ 9	+19	- 7	+ 9	+14	+ 3	+18	+10	-12
K	+ 7	+ 1	-10	+18	+ 8	- 8	- 6	+ 1	+ 2	- 3	+ 1
L	- 5	+ 4	-10	+12	- 9	+10	-17	- 5	- 2	+ 5	- 8
M	+ 1	+ 8	- 7	+19	+10	-11	-10	- 3	+ 8	- 3	+ 1
N	+ 9	+ 5	+ 3	- 2	+ 7	+ 8	-16	+ 3	- 3	-20	- 4
O	- 8	-13	-11	+ 9	+ 1	- 7	+ 1	+ 2	- 4	+ 9	- 2

[1] Rounded to nearest whole number

[2] Algebraic Sum of Deviations: Indicative of Direction

APPENDIX D-3

FIFTEEN TEACHERS CLASSIFIED INTO FIVE PANELS OF JUDGES BY DIRECTIONAL FACTOR

	Panel 1	Panel 2	Panel 3	Panel 4	Panel 5
Lowest One-Third (Minus Direction)	Tchr. D -96	Tchr. A -55	Tchr. I -25	Tchr. L -25	Tchr. O -23
Highest One-Third (Plus Direction)	Tchr. F +84	Tchr. E +58	Tchr. G +42	Tchr. J +31	Tchr. B +20
Middle One-Third (Plus & Minus Direction)	Tchr. C +19	Tchr. K +11	Tchr. N -10	Tchr. H + 7	Tchr. M +13
Cumulative Direction of Each Panel	+ 7	+14	+ 7	+13	+10
Deviation of Each Panel from Mean (+10.2) Cumulative Direction	-3.2	+3.8	-3.2	+2.8	-.2

Average Deviation of Panels from Mean Cumulative Direction: 2.64

Mean Cumulative Direction per Theme per Panel: +.92 points (on 100 point scale)

APPENDIX E-1

Students' Reactions to Use of Recorded Comments

CODE:
Y: yes
N: no
DK: don't know
PS: perhaps so
PN: probably not
R: respondents (number)

Items	R	% Y	% N	% DK	% PS	% PN
1. I got more out of freshman English because the records were used to comment on my themes.	277	44	21	9	21	5*
2. The teacher's recorded comments eased my task of revising papers.	276	60	24	6	6	4
3. Of the four categories--content, organization, sentences, mechanics--the recorded comments helped me most to improve my content.	245	31	33	7	23	6
4. If I had a choice between the recorded comments on my papers and a personal conference with my instructor, I would choose the personal conference.	263	62	16	6	10	6
5. I would like to have records used for teachers' comments on my papers in future English classes.	205	55	23	8	11	3

*All percentages rounded off.

		% R	% Y	% N	% DK	% PS	% PN
6.	The recorded comments on a given paper I found to be of real help when I started writing another paper.	266	46	21	11	20	2
7.	Of the four categories--content, organization, sentences, mechanics-- the recorded comments helped me most to improve my organization.	262	27	38	8	22	5
8.	If I had a choice between the recorded comments on my papers and the same comments written out and attached to my paper, I would choose the comments written out and attached to my paper.	267	33	48	5	29	5
9.	I wish I had had recorded comments on the freshmen English papers that I wrote last semester.	270	56	27	5	11	1
10.	I liked being able to play the record over again so that I could double-check my mistakes.	251	66	17	4	12	1
11.	Of the four categories--content, organization, sentences, mechanics--the recorded comments helped me most to improve my sentences.	259	20	42	8	21	9
12.	Because of the recorded comments, I received lower grades on my papers than I probably would have if my teacher had used the traditional method.	256	5	57	23	5	10
13.	I thought the records were more effective at the beginning of the semester than toward the end.	264	21	48	10	13	8

	% R	% Y	% N	% DK	% PS	% PN
14. After having completed a revision to the teacher's satisfaction, I saved the record to play back later on as a check on whether I continued making the same mistakes.	253	34	57	2	4	3
15. Of the four categories--content, organization, sentences, mechanics--the recorded comments helped me most to improve my mechanics.	261	25	38	7	22	8
16. With the recorded comments I believe my teacher spent too much time arguing with my ideas rather than talking about my grammar and punctuation and spelling.	269	12	73	1	8	6
17. I thought the records were more effective toward the end of the semester than at the beginning.	259	22	47	13	10	8
18. The recorded comments confused me more than they straightened out my thinking.	258	12	75	5	5	3
19. Of the two categories--content and organization versus sentences and mechanics--the recorded comments helped me most to improve my content and organization.	244	50	32	11	6	1
20. I liked the use of records from the beginning.	274	52	31	7	9	1
21. Because of the recorded comments, I received higher grades on my papers than I probably would have if my teacher had used the traditional method.	263	15	34	24	17	10

22.	I learned more about how to write a paper this semester from the recorded comments than I did last semester with the traditional method.	257	40	27	11	15	7
23.	I never liked the records at all.	247	13	72	4	7	4
24.	Of the two categories--content and organization versus sentences and mechanics--the recorded comments helped me most to improve my sentences and mechanics.	265	29	44	6	11	10

SAMPLE

Student Questionnaire

To: Students whose themes were graded with records

Directions: For each statement made below, check the one answer true for you. If you do not choose to answer a particular item, do not check any blanks.

1. I got more out of freshman English because the records were used to comment on my themes.
yes__; no__; don't know__; perhaps so__; probably not__

2. The teacher's . . .

APPENDIX E-2

Students' Attitudes Toward Records Categorized

a. Effect of Recorded Comment on Improvement in Four Main Criteria for Scoring Themes

		Percentage of Replies		
		favorable	unfavorable	don't know
3.	Disks improved content most	54	39	7
7.	Disks improved organization most	49	43	8
11.	Disks improved sentences most	41	51	8
15.	Disks improved mechanics most	47	46	7
19.	Disks improved content and organization most	56	33	11
24.	Disks improved sentences and mechanics most	40	54	6

b. Students' Like or Dislike of Records

		Percentage of Replies		
		favorable	unfavorable	don't know
5.	I want disks used in future English classes	66	26	8
20.	I always liked records	61	32	7
23.	I never liked records	20	76*	4
13.	Disks were more effective at beginning	34	56	10
17.	Disks were more effective toward end	32	55	13

*Actually favorable to disks because of wording of item

c. Effect of Recorded Comment on Revision

Items Summarized	Percentage of Replies		
	favor- able	unfavor- able	don't know
2. Disks helped my revisions	66	28	6
18. Disks confused me	17	78*	5
14. I saved disks for playing later on	38	60	2
10. I played disks back as re-check of revision	78	18	4

*Actually favorable to disks because of wording of item

d. Effect of Recorded Comments on Subsequent Papers

Items Summarized	Percentage of Replies		
	favor- able	unfavor- able	don't know
6. Disks helped me to write new themes	66	23	11
16. Teachers argued too much about ideas	20	79*	1

*Actually favorable to use of disks for criticizing ideas rather than mechanics

e. Effect of Recorded Comments on Learning in General

Items Summarized	Percentage of Replies		
	favorable	unfavorable	don't know
12. Disks lowered my grades	10	67*	23
21. Disks raised my grades	32	44	24
22. Disks taught me better how to write	55	34	11
1. Disks taught me more in freshmen English	65	26	9
9. I wish disks had been used in past	67	28	5
5. I want disks used in future	66	26	8
8. I prefer written comments to disks	42	53*	5
4. I prefer disks to personal conference	22	72	6

*Actually favorable to use of disks because of wording of item

APPENDIX F-1

Fifteen Teachers' Reactions to Use of Recorded Comments

Y: yes
N: no
DK: don't know
PS: perhaps so
PN: probably not

Items	% Y	% N	% DK	% PS	% PN
1. I'm afraid that the use of recorded commentaries is but another step toward our dehumanizing education.	0	87	0	7	7
2. My impression is that students in the experimental section revised their papers in a manner superior to that of those in the control section.	33	47	0	7	13
3. Of the four categories--content, organization, sentences, mechanics--I think my recorded comments helped students most in improving their content.	27	33	7	20	13
4. My grading of themes in the experimental section was superior to that of the control section.	7	40	20	20	13
5. Using the records, I think I gave lower grades than I would have otherwise.	0	100	0	0	0
6. Of the four categories--content, organization, sentences, mechanics--I think my recorded comments helped students most in improving their organization.	27	27	7	33	7

		% Y	% N	% DK	% PS	% PN
7.	Using the recorded commentary, I felt that I was personally aiding a student more than I do with the traditional method.	53	7	7	20	13
8.	In the future, I should like to use the recorded commentary in other freshman English classes I teach.	47	20	7	27	0
9.	Of the four categories--content, organization, sentences, mechanics--I think my recorded comments helped students most in improving their sentences.	27	27	13	13	20
10.	I never liked the use of the records at all, from beginning to end.	7	93	0	0	0
11.	Using the records, I think I gave higher grades than I would have otherwise.	20	73	0	0	7
12.	I think the experiment was too short to prove anything of value.	7	40	13	20	20
13.	Of the four categories--content, organization, sentences, mechanics--I think my recorded commentary helped students most in improving their mechanics.	0	67	7	7	20
14.	I think the experiment was a success.	47	0	40	7	7
15.	Throughout the semester I always felt a little silly about cutting a record--as though there were something unprofessional and gimmicky about the whole procedure.	7	93	0	0	0
16.	For the value that students receive from the recorded commentary, cutting a record takes too much of the instructor's time.	47	20	27	7	0
17.	I think my grading was unaffected by using records; that is, my grades were no higher or no lower than otherwise.	73	27	0	0	0
18.	Of the two categories--content and organization versus sentences and mechanics--I think my recorded comments helped students most in improving their content and organization.	47	13	13	20	7

	% Y	% N	% DK	% PS	% PN
19. In the future, I should like to use the recorded commentary for all papers written in English classes: from sophomore level through graduate courses.	33	47	7	0	13
20. I found that checking a student's revisions was more difficult with the recorded commentary than with the traditional method.	40	33	7	7	13
21. I think the experiment was a failure.	0	33	27	7	7
22. I thought my commentaries were as good at the beginning of the semester as they were toward the end.	27	20	13	20	20
23. Even though the students in the experimental sections might not show any statistically significant improvement as compared to those in the control sections, I would still recommend the recorded commentary as superior to the traditional method of reading themes.	27	13	20	20	20
24. Being confronted with a blank record, I felt compelled to talk--if no more than to give the student his "money's worth"--even when I didn't have much of value to say.	13	73	0	13	0
25. Of the two categories--content and organization versus sentences and mechanics--I think my recorded comments helped students most in improving their sentences and mechanics.	13	60	7	7	13
26. I'd recommend the recorded commentary to any teacher who has papers to grade--regardless of department.	27	40	0	20	13
27. Even though the students in the experimental sections might show statistically significant improvement as compared to those in control sections, I would still recommend the traditional method of grading themes over the experimental method.	7	40	27	7	20

	% Y	% N	% DK	% PS	% PN
28. As the semester progressed, I thought that my commentaries improved.	13	27	7	47	7

S A M P L E

Teacher Questionnaire

To: Teachers of experimental sections of freshman English

Directions: For each statement made below, check the one answer that is true for you.

1. I'm afraid that the use of recorded commentaries is but another step toward our dehumanizing education.
   yes _____; no ______; don't know _______; perhaps so _____; probably not ______.

2. My impression is that. . .

APPENDIX F-2

Fifteen Teachers' Attitudes Toward Records Categorized

a. Effect of Recorded Comment on Students' Improvement in Four Main Criteria for Scoring Themes

Items Summarized	Percentage of Replies		
	favorable	unfavorable	don't know
3. Disks improved students' content most	47	47	6
6. Disks improved students' organization most	60	33	7
9. Disks improved students' sentences most	40	47	13
13. Disks improved students' mechanics most	7	87	7
18. Disks improved students' content and organization most	67	20	13
25. Disks improved students' sentences and mechanics most	20	73	7

b. Teachers' Like or Dislike of Records

	Items Summarized	Percentage of Replies		
		favorable	unfavorable	don't know
1.	Disks dehumanize education	7	93*	0
10.	I never liked the disks	7	93*	0
12.	Experiment with disks was too short	27	60	13
14.	Experiment was successful	53	7	40
21.	Experiment was a failure	7	7	40
15.	Cutting disks was unprofessional	7	93*	0
24.	I felt compulsion to fill disk	27	73*	0

*Actually favorable to use of disks because of wording of item.

c. Effect of Recorded Comment on Students' Revisions

	Items Summarized	Percentage of Replies		
		favorable	unfavorable	don't know
2.	Records resulted in superior revisions	40	60	0
20.	Records made checking revision difficult	47	47	6

d. Teachers' Attitude Toward Recording as a Technique

	Items Summarized	Percentage of Replies		
		favorable	unfavorable	don't know
28.	Commentaries improved as semester progressed	60	33	7
22.	Commentaries were as good at beginning as at end	47	40	13
16.	Cutting record takes too much of teacher's time	53*	20	27
7.	Cutting record gives teacher feeling of giving student more personal attention	73	20	7

*Actually unfavorable toward disks because of wording of item

e. Effect of Cutting Records on Teachers' Grading

	Items Summarized	Percentage of Replies		
		favorable	unfavorable	don't know
4.	Grading with records superior	27	53	20
5.	Grades lower because of records	0	100*	0
17.	Grading unaffected by use of records	73	27	0
11.	Grades higher because of records	20	80*	0

*Actually favorable to use of disks because of wording of item

f. Teachers' Attitude Toward Future Use of Recorded Comments

	Items Summarized	Percentage of Replies		
		favorable	unfavorable	don't know
8.	Want records in future freshman English classes	73	20	7
19.	Want records used in more advanced English classes	33	60	7
26.	Recommend records for use in any department	47	53	0
27.	Recommend traditional method even though experiment statistically favorable to records	13	60*	27
23.	Recommend recorded commentary even though experiment not statistically favorable to records	47	33	20

*Actually favorable to use of disks because of wording of item.

APPENDIX G-1

Correlation Coefficients
of
Judges Paired Within Panels
derived from
Pre-Theme Scores and Pre-Error Counts
for
Experimental and Control Groups

Judges[1]	CO[2]	SM[3]	Err[4]	COSM[5]
11, 12	.42	.60	.70	.55
11, 13	.55	.60	.73	.60
12, 13	.54	.62	.66	.61
21, 22	.62	.71	.84	.71
21, 23	.12	.54	.75	.30
22, 23	.03	.67	.78	.37
31, 32	.34	.48	.57	.43
31, 33	.55	.61	.60	.59
32, 33	.34	.47	.78	.39
41, 42	.31	.25	.74	.26
41, 43	.33	.38	.67	.38
42, 43	.37	.60	.70	.48
51, 52	.52	.62	.74	.64
51, 53	.57	.73	.70	.73
52, 53	.39	.39	.65	.52

[1] First digit identifies panel; second digit, judge

[2] Content and organization: 30 points each

[3] Sentences and mechanics: 20 points each

[4] Actual count of errors in ten categories

[5] Content, organization, sentences, and mechanics

APPENDIX G-2

Mean Scores of STEP

(All Sections, Regular Sections, Superior Sections)

		Experimental Group			Control Group		
	Panel	Pre-STEP	Post-STEP	Difference	Pre-STEP	Post-STEP	Difference
All Secs	1	306.51	312.31	+5.80 (+1.6%)	308.38	312.06	+3.68 (+1.2%)
	2	307.67	312.23	+4.56 (+1.5%)	305.02	314.35	+9.33 (+3.1%)
	3	301.05	304.53	+3.48 (+1.2%)	302.53	304.18	+1.65 (+0.5%)
	4	303.27	306.93	+3.66 (+1.2%)	302.05	305.35	+3.30 (+1.1%)
	5	302.71	312.06	+9.35 (+3.1%)	298.73	305.88	+7.15 (+2.4%)
				Mean: +1.7%			Mean: +1.7%
Regular Secs	1	299.62	307.41	+7.79 (+2.6%)	301.60	306.68	+5.08 (+1.7%)
	2	300.38	308.53	+8.15 (+2.7%)	299.28	311.80	+12.52 (+4.2%)
	3	301.05	304.53	+3.48 (+1.2%)	302.53	304.18	+1.65 (+0.5%)
	4	303.27	306.93	+3.66 (+1.2%)	302.05	305.35	+3.30 (+1.1%)
	5	302.71	312.06	+9.35 (+3.1%)	298.73	305.88	+7.15 (+2.4%)
				Mean: +2.2%			Mean: +2.0%
Superior Secs	1	316.23	319.23	+3.00 (+0.9%)	316.09	318.18	+2.09 (+0.7%)
	2	317.15	317.05	-0.10 (0%)	313.47	318.11	+4.64 (+1.5%)
				Mean: +0.5%			Mean: +1.1%

APPENDIX H-1

Mean Scores of COSM for All Students

(All Sections, Regular Sections, Superior Sections)

		Experimental Group			Control Group		
	Panel	Pre-Theme	Post-Theme	Difference	Pre-Theme	Post-Theme	Difference
All Secs	1	64.96	67.11	+2.15 (+3.3%)	70.25	69.71	-0.54 (-0.7%)
	2	70.83	73.86	+3.03 (+4.2%)	68.87	73.14	+4.27 (+6.2%)
	3	66.08	69.91	+3.83 (+5.8%)	68.07	69.99	+1.92 (+2.8%)
	4	70.41	68.28	-2.13 (-3.0%)	68.71	69.03	+0.32 (+0.4%)
	5	70.36	72.53	+2.17 (+3.1%)	66.16	68.91	+2.75 (+4.1%)
				Mean: +2.7%			Mean: +2.3%
Regular Secs	1	63.82	63.67	-0.15 (-0.2%)	68.61	66.50	-2.11 (-3.1%)
	2	67.96	72.56	+4.60 (+6.7%)	65.21	71.57	+6.36 (+9.7%)
	3	66.08	69.91	+3.83 (+5.8%)	68.07	69.99	+1.92 (+2.8%)
	4	70.41	68.28	-2.13 (-3.0%)	68.71	69.03	+0.32 (+0.4%)
	5	70.36	72.53	+2.17 (+3.1%)	66.16	68.91	+2.75 (+4.1%)
				Mean: +2.5%			Mean: +2.8%
Superior Secs	1	66.60	70.54	+3.94 (+5.9%)	72.12	73.51	+1.39 (+1.9%)
	2	74.57	75.55	+0.98 (+1.3%)	74.25	78.78	+4.53 (+6.1%)
				Mean: +3.6%			Mean: +4.0%

APPENDIX H-2

Mean Scores of COSM for Paired Students

(All Sections, Regular Sections, Superior Sections)

		Experimental Group			Control Group		
	Panel	Pre-Theme	Post-Theme	Difference	Pre-Theme	Post-Theme	Difference
All Secs	1	64.49	68.17	+3.68 (+5.7%)	68.16	69.60	+1.44 (+2.1%)
	2	70.70	73.90	+3.20 (+4.5%)	69.56	74.08	+4.52 (+6.5%)
	3	67.85	70.19	+2.34 (+3.4%)	67.49	68.88	+1.39 (+2.1%)
	4	70.35	68.55	-1.98 (-2.8%)	68.17	68.16	-0.01 (0 %)
	5	68.63	71.06	+2.43 (+3.5%)	65.85	69.17	+3.32 (+5.1%)
				Mean: +2.9%)			Mean: +3.3%
Regular Secs	1	60.71	62.82	+2.11 (+3.4%)	65.75	65.79	+0.04 (0 %)
	2	67.98	72.27	+4.29 (+6.2%)	65.39	72.52	+7.13 (+10.9%)
	3	67.85	70.19	+2.34 (+3.4%)	67.49	68.88	+1.39 (+2.1%)
	4	70.53	68.55	-1.98 (-2.8%)	68.17	68.16	-0.01 (0 %)
	5	68.63	71.06	+2.43 (+3.5%)	65.85	69.17	+3.32 (+5.1%)
				Mean: +2.7%			Mean: +3.6%
Superior Secs	1	68.55	73.92	+5.37 (+7.8%)	70.73	73.68	+2.95 (+4.1%)
	2	74.21	76.00	+1.79 (+2.4%)	74.92	76.09	+1.17 (+1.6%)
				Mean: +5.1%			Mean: +2.9%

APPENDIX H-3

Mean Scores of CO for All Students

(All Sections, Regular Sections, Superior Sections)

		Experimental Group			Control Group		
	Panel	Pre-Theme	Post-Theme	Difference	Pre-Theme	Post-Theme	Difference
All Secs	1	38.40	40.71	+2.31 (+6.0%)	42.02	41.94	-0.08 (0.0%)
	2	41.10	43.97	+2.87 (+6.9%)	40.02	43.75	+3.77 (+9.4%)
	3	40.07	42.69	+2.62 (+6.5%)	40.52	42.35	+1.83 (+4.5%)
	4	42.01	41.57	-0.44 (-1.0%)	42.31	41.66	-0.65 (-1.5%)
	5	42.90	44.28	+1.38 (+3.2%)	40.81	43.01	+2.20 (+5.4%)
				Mean: +4.3%			Mean: +3.7%
Regular Secs	1	38.08	39.50	+1.42 (+3.7%)	41.00	40.87	-0.13 (-0.3%)
	2	39.03	43.40	+4.10 (+10.4%)	38.25	43.40	+5.15 (+13.4%)
	3	40.07	42.69	+2.62 (+6.5%)	40.52	42.35	+1.83 (+4.5%)
	4	42.01	41.57	-0.44 (-1.0%)	42.31	41.66	-0.65 (-1.5%)
	5	42.90	44.28	+1.38 (+3.2%)	40.81	43.01	+2.20 (+5.4%)
				Mean: +4.6%			Mean: +4.3%
Superior Secs	1	38.86	42.43	+3.57 (+9.2%)	43.18	43.21	+0.03 (0.0%)
	2	43.43	44.72	+1.29 (+2.9%)	42.64	44.29	+1.65 (+3.9%)
				Mean: +6.1%			Mean: +2.0%

APPENDIX H-4

Mean Scores of CO for Paired Students

(All Sections, Regular Sections, Superior Sections)

		Experimental Group			Control Group		
	Panel	Pre-Theme	Post-Theme	Difference	Pre-Theme	Post-Theme	Difference
All Secs	1	38.06	41.36	+3.30 (+8.6%)	40.57	42.27	+1.70 (+4.2%)
	2	41.06	47.43	+6.37 (+15.5%)	40.13	44.21	+4.08 (+10.2%)
	3	41.06	42.85	+1.79 (+4.3%)	39.99	41.77	+1.78 (+4.5%)
	4	42.32	41.34	-0.98 (-2.3%)	41.87	41.09	-0.78 (-1.9%)
	5	41.66	43.37	+1.71 (+4.3%)	39.97	42.69	+2.72 (+6.8%)
				Mean: +6.1%			Mean: +4.8%
Regular Secs	1	36.33	39.07	+2.74 (+7.5%)	39.13	41.17	+2.04 (+5.2%)
	2	39.42	43.20	+3.78 (+9.5%)	38.03	44.01	+5.98 (+15.6%)
	3	41.06	42.85	+1.79 (+4.3%)	39.99	41.77	+1.78 (+4.5%)
	4	42.32	41.34	-0.98 (-2.3%)	41.87	41.09	-0.78 (-1.9%)
	5	41.66	43.37	+1.71 (+4.3%)	39.97	42.69	+2.72 (+6.8%)
				Mean: +4.7%			Mean: +6.0%
Superior Secs	1	39.93	44.52	+4.59 (+11.5%)	42.11	43.45	+1.34 (+3.2%)
	2	43.16	45.26	+2.10 (+4.9%)	42.83	44.47	+1.64 (+3.8%)
				Mean: +8.2%			Mean: +3.5%

APPENDIX H-5

Mean Scores of SM for All Students

(All Sections, Regular Sections, Superior Sections)

		Experimental Group			Control Group		
	Panel	Pre-Theme	Post-Theme	Difference	Pre-Theme	Post-Theme	Difference
All Secs	1	26.56	26.39	-0.17 (-0.7%)	28.23	27.77	-0.46 (-1.6%)
	2	29.73	29.88	+0.15 (+0.5%)	28.83	29.38	+0.55 (+1.9%)
	3	26.07	27.25	+1.18 (+4.5%)	27.64	27.63	-0.01 (0.0%)
	4	28.41	26.70	-1.71 (-6.0%)	26.39	27.36	+0.97 (+3.6%)
	5	27.45	28.24	+0.79 (+2.8%)	25.35	25.90	+0.55 (+2.2%)
				Mean: +0.2%			Mean: +1.2%
Regular Secs	1	25.73	25.17	-0.56 (-2.2%)	27.61	25.62	-1.99 (-7.2%)
	2	28.65	29.15	+0.50 (+1.7%)	26.96	28.17	+1.21 (+4.1%)
	3	26.07	27.25	+1.18 (+4.5%)	27.64	27.63	-0.01 (0.0%)
	4	28.41	26.70	-1.71 (-6.0%	26.39	27.36	+0.97 (+3.6%)
	5	27.45	28.24	+0.79 (+2.8%)	25.35	25.90	+0.55 (+2.2%)
				Mean: +0.2%			Mean: +0.5%
Superior Secs	1	27.74	28.11	+0.37 (+1.3%)	28.94	30.30	+1.36 (+4.6%)
	2	31.13	30.83	-0.30 (-0.9%)	31.60	31.15	-0.45 (-1.4%)
				Mean: +0.2%			Mean: +1.6%

APPENDIX H-6

Mean Scores of SM for Paired Students

(All Sections, Regular Sections, Superior Sections)

		Experimental Group			Control Group		
	Panel	Pre-Theme	Post-Theme	Difference	Pre-Theme	Post-Theme	Difference
All Secs	1	26.42	26.48	+0.06 (+0.2%)	27.58	26.99	-0.59 (-2.1%)
	2	29.61	29.80	+0.19 (+0.6%)	29.42	29.86	+0.44 (+1.5%)
	3	26.79	27.34	+0.55 (+2.1%)	27.49	27.10	-0.39 (-1.4%)
	4	28.17	27.20	-0.97 (-3.4%)	26.30	27.04	+0.74 (+2.8%)
	5	26.30	27.68	+1.38 (+5.2%)	25.88	26.47	+0.59 (+2.2%)
				Mean: +0.9%			Mean: +0.6%
Regular Secs	1	24.37	23.75	-0.62 (-2.5%)	26.62	24.62	-2.00 (-7.5%)
	2	28.55	29.07	+0.52 (+1.8%)	27.35	28.49	+1.14 (+4.2%)
	3	26.79	27.34	+0.55 (+2.1%)	27.49	27.10	-0.39 (-1.4%)
	4	28.17	27.20	-0.97 (-3.4%)	26.30	27.04	+0.74 (+2.8%)
	5	26.30	27.68	+1.38 (+5.2%)	25.88	26.47	+0.59 (+2.2%)
				Mean: +0.6%			Mean: +0.1%
Superior Secs	1	28.62	29.40	+0.78 (+2.7%)	28.62	30.23	+1.61 (+5.6%)
	2	31.04	30.73	-0.31 (-0.9%)	32.09	31.61	-0.48 (-1.5%)
				Mean: +0.9%			Mean: +2.0%

APPENDIX J-1

Mean Error-Counts (Ten Categories) for All Students

Sent Sense, Agr, Ref, Vb, Case, Mod, Paral, Sp, Cap, Pn

All Sections, Regular Sections, Superior Sections

		Experimental Group			Control Group		
	Panel	Pre-Theme	Post-Theme	Difference	Pre-Theme	Post-Theme	Difference
All Secs	1	8.24	9.58	-1.29 (-14.4%)	7.67	8.21	-0.54 (- 7.0%)
	2	5.74	6.55	-0.81 (-14.1%)	6.47	6.75	-0.28 (- 4.3%)
	3	7.31	5.84	+1.47 (+20.1%)	5.18	5.25	-0.07 (- 1.3%)
	4	6.03	7.40	-1.37 (-22.7%)	8.46	7.02	+1.44 (+17.0%)
	5	7.49	7.11	+0.38 (+ 5.0%)	8.37	8.49	-0.12 (- 1.4%)
				Mean: - 5.2%			Mean: + 0.6%
Regular Secs	1	9.16	10.49	-1.33 (-14.5%)	9.23	9.76	-0.53 (- 5.8%)
	2	6.74	7.65	-0.91 (-13.5%)	8.03	7.97	+0.06 (+ 0.7%)
	3	7.31	5.84	+1.47 (+20.1%)	5.18	5.25	-0.07 (- 1.3%)
	4	6.03	7.40	-1.37 (-22.7%)	8.46	7.02	+1.44 (+17.0%)
	5	7.49	7.11	+0.38 (+ 5.0%)	8.37	8.49	-0.12 (- 1.4%)
				Mean: - 5.1%			Mean: + 1.8%
Superior Secs	1	7.01	8.25	-1.24 (-17.7%)	5.90	6.45	-0.55 (- 9.3%)
	2	4.36	5.18	-0.82 (-18.8%)	4.19	4.96	-0.77 (-18.3%)
				Mean: -18.2%			Mean: -13.8%

APPENDIX J-2

Mean Error-Counts (Ten Categories) for Paired Students

Sent Sense, Agr, Ref, Vb, Case, Mod, Paral, Sp, Cap, Pn

All Sections, Regular Sections, Superior Sections

		Experimental Group			Control Group		
	Panel	Pre-Theme	Post-Theme	Difference	Pre-Theme	Post-Theme	Difference
All Secs	1	8.82	9.68	-0.86 (- 9.7%)	8.25	8.86	-0.61 (-7.4%)
	2	5.76	6.40	-0.64 (-11.1%)	5.63	5.86	-0.23 (-4.1%)
	3	6.24	5.76	+0.48 (+ 7.7%)	5.24	5.35	-0.09 (-1.7%)
	4	5.99	7.15	-1.16 (-19.3%)	8.50	7.07	+1.43 (+16.8%)
	5	7.84	7.18	+0.66 (+ 8.4%)	7.44	7.87	-0.43 (-5.8%)
				Mean: - 4.8%			Mean: -0.4%
Regular	1	10.18	12.19	-2.01 (-19.7%)	10.35	11.21	-0.86 (-8.3%)
	2	6.81	7.33	-0.52 (- 7.6%)	6.82	6.87	-0.05 (-0.8%)
	3	6.24	5.76	+0.48 (+ 7.7%)	5.24	5.35	-0.09 (-1.7%)
	4	5.99	7.15	-1.16 (-19.3%)	8.50	7.07	+1.43 (+16.8%)
	5	7.84	7.18	+0.66 (+ 8.4%)	7.44	7.87	-0.43 (-5.8%)
				Mean: - 7.1%			Mean: +0.4%
Superior	1	9.20	7.16	+0.04 (+ 0.6%)	5.81	6.52	-0.71 (-12.2%)
	2	4.40	5.33	-0.93 (-21.1%)	4.04	4.71	-0.67 (-16.6%
				Mean: -10.3%			Mean: -14.4%

Appendix J-3

Mean Error-Counts (Seven Categories) for All Students

Sent Sense, Agr, Ref, Vb, Case, Mod, Paral

All Sections, Regular Sections, Superior Sections

		Experimental Group			Control Group		
	Panel	Pre-Theme	Post-Theme	Difference	Pre-Theme	Post-Theme	Difference
All Secs	1	3.11	3.25	-0.14 (- 4.5%)	3.11	2.96	+0.15 (+4.8%)
	2	1.77	1.81	-0.04 (- 2.3%)	1.99	2.19	-0.20 (-10.1%)
	3	2.27	1.90	+0.37 (+16.3%)	1.89	1.49	+0.40 (+21.2%)
	4	1.96	2.22	-0.26 (-13.3%)	2.53	2.34	+0.19 (+7.5%)
	5	3.04	3.01	+0.03 (+ 0.9%)	2.85	2.72	+0.13 (+4.6%)
				Mean: - 0.6%			Mean: +5.6%
Regular Secs	1	3.40	3.15	+0.25 (+ 7.4%)	3.93	3.20	+0.73 (+18.7%)
	2	1.81	1.95	-0.14 (- 7.7%)	2.45	2.75	-0.30 (-12.2%)
	3	2.27	1.90	+0.37 (+16.3%)	1.89	1.49	+0.40 (+21.2%)
	4	1.96	2.22	-0.26 (-13.3%)	2.53	2.34	+0.19 (+ 7.5%)
	5	3.04	3.01	+0.03 (+ 0.9%)	2.85	2.72	+0.13 (+ 4.6%)
				Mean: + 0.7%			Mean: + 8.2%
Superior Secs	1	2.72	3.41	-0.69 (-25.4%)	2.18	2.69	-0.51 (-23.4%)
	2	1.71	1.65	+0.06 (+ 3.5%)	1.31	1.37	-0.06 (- 4.6%)
				Mean: -11.0%			Mean: -14.0%

APPENDIX J-4

Mean Error-Counts (Seven Categories) for Paired Students

Sent Sense, Agr, Ref, Vb, Case, Mod, Paral

All Sections, Regular Sections, Superior Sections

		Experimental Group			Control Group		
	Panel	Pre-Theme	Post-Theme	Difference	Pre-Theme	Post-Theme	Difference
All Secs	1	2.88	3.13	-0.25 (- 8.6%)	3.23	2.39	+0.84 (+26.0%)
	2	1.82	1.82	0 (0.0%)	1.81	1.99	-0.18 (- 9.9%)
	3	2.01	1.94	+0.07 (+ 3.5%)	1.79	1.55	+0.24 (+13.4%)
	4	1.99	2.20	-0.21 (-10.6%)	2.54	2.30	+0.24 (+ 9.4%)
	5	3.10	3.03	+0.07 (+ 2.3%)	2.53	2.53	0 (0)
				Mean: - 2.7%			Mean: + 7.8%
Regular Secs	1	3.31	3.33	-0.02 (- 0.6%)	4.31	3.09	+1.22 (+28.3%)
	2	1.88	1.64	+0.24 (+12.8%)	2.33	2.41	-0.08 (- 3.4%)
	3	2.01	1.94	+0.07 (+ 3.5%)	1.79	1.55	+0.24 (+13.4%)
	4	1.99	2.20	-0.21 (-10.6%)	2.54	2.30	+0.24 (+ 9.4%)
	5	3.10	3.03	+0.07 (+ 2.3%)	2.53	2.53	0 (0)
				Mean: - 2.7%			Mean: + 9.5%
Superior Secs	1	2.38	2.93	-0.55 (-23.5%)	1.99	2.90	-0.91 (-45.7%)
	2	1.78	2.02	-0.24 (-13.5%)	1.14	1.49	-0.35 (-30.7%)
				Mean: -18.5%			Mean: -38.2%

APPENDIX J-5

Mean Error-Counts (Three Categories) for All Students

Sp, Cap, Pn

All Sections, Regular Sections, Superior Sections

		Experimental Group			Control Group		
	Panel	Pre-Theme	Post-Theme	Difference	Pre-Theme	Post-Theme	Difference
All Secs	1	5.16	6.33	-1.17 (-22.7%)	4.56	5.24	-0.68 (-14.9%)
	2	3.97	4.86	-0.89 (-22.4%)	4.48	4.56	-0.06 (- 1.3%)
	3	5.04	3.93	+1.11 (+22.0%)	3.28	3.75	-0.47 (-14.3%)
	4	4.04	5.18	-1.14 (-28.2%)	5.93	4.67	+1.26 (+21.2%)
	5	4.46	4.10	+0.36 (+ 8.1%)	5.52	5.77	-0.25 (- 4.5%)
				Mean: - 8.6%			Mean: - 2.8%
Regular Secs	1	5.76	7.35	-1.59 (-27.6%)	5.29	6.56	-1.27 (-24.0%)
	2	4.90	5.71	-0.81 (-16.5%)	5.54	5.23	+0.31 (+ 5.6%)
	3	5.04	3.93	+1.11 (+22.0%)	3.28	3.75	-0.47 (-14.3%)
	4	4.04	5.18	-1.14 (-28.2%)	5.93	4.67	+1.26 (+21.2%)
	5	4.46	4.10	+0.36 (+ 8.1%)	5.52	5.77	-0.25 (- 4.5%)
				Mean: - 8.4%			Mean: - 3.3%
Superior Secs	1	4.29	4.84	-0.55 (-12.8%)	3.72	3.75	-0.03 (- 0.8%)
	2	2.64	3.53	-0.89 (-37.5%)	2.88	3.58	-0.70 (-24.3%)
				Mean: -25.2%			Mean: -12.6%

APPENDIX J-6

Mean Error-Counts (Three Categories) for Paired Students

Sp, Cap, Pn

All Sections, Regular Sections, Superior Sections

		Experimental Group			Control Group		
	Panel	Pre-Theme	Post-Theme	Difference	Pre-Theme	Post-Theme	Difference
All Secs	1	5.60	6.54	-0.94 (-16.8%)	5.01	6.03	-1.02 (-23.6%)
	2	3.93	4.64	-0.71 (-18.1%)	3.82	3.87	-0.05 (- 1.3%)
	3	4.23	3.82	+0.41 (+ 9.7%)	3.45	3.80	-0.35 (-10.1%)
	4	3.97	4.94	-0.97 (-24.4%)	5.29	4.76	+1.47 (+27.8%)
	5	4.73	4.14	+0.59 (+12.5%)	4.90	5.34	-0.44 (+ 8.9%)
				Mean: - 7.7%			Mean: + 0.3%
Regular Secs	1	6.86	8.85	-1.99 (-28.9%)	6.04	8.12	-2.08 (-34.4%)
	2	4.96	5.68	-0.72 (-14.5%)	4.53	4.45	+0.08 (+ 1.7%)
	3	4.23	3.82	+0.41 (+ 9.7%)	3.45	3.80	-0.35 (-10.1%)
	4	3.97	4.94	-0.97 (-24.4%)	5.29	4.76	+1.47 (+27.8%)
	5	4.73	4.14	+0.59 (+12.5%)	4.90	5.34	-0.44 (+ 8.9%)
				Mean: - 9.1%			Mean: - 1.2%
Superior Secs	1	4.15	4.23	-0.08 (- 1.9%)	3.81	3.95	-0.14 (- 3.7%)
	2	2.62	3.31	-0.69 (-26.3%)	2.90	3.21	-0.31 (-10.3%)
				Mean: -14.1%			Mean: - 7.0%